AD NUMBER: AD0853356

LIMITATION CHANGES

TO:

Approved for public release; distribution is unlimited.

FROM:

Distribution authorized to U.S. Gov't. agencies and their contractors; Export Control; 1 Jan 1969. Other requests shall be referred to the Office of Naval Research, Field Projects Branch, Washington, D.C., 20360.

AUTHORITY

ONR LTR, 28 JUL 1977

SU-SEL-69-007

Methods and Applications of Computer Raytracing

by

Thomas A. Croft

January 1969

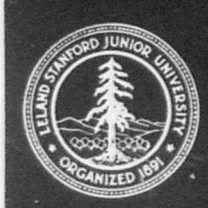
Technical Report No. 112

Prepared under
Office of Naval Research Contract
Nonr-225(64), NR 088 019, and
Advanced Research Projects Agency ARPA Order 196

RADIOSCIENCE LABORATORY

STANFORD ELECTRONICS LABORATORIES

STANFORD UNIVERSITY · STANFORD, CALIFORNIA



METHODS AND APPLICATIONS OF COMPUTER RAYTRACING

by

Thomas A. Croft

January 1969

Technical Report No. 112

Prepared under
Office of Naval Research Contract
Nonr-225(64), NR 088-019, and
Advanced Research Projects Agency
ARPA Order No. 196

Radioscience Laboratory
Stanford Electronics Laboratories
Stanford University Stanford, California

ABSTRACT

For about five years, considerable effort was devoted at this laboratory to the exploitation of the digital computer as a means for simulating ionospheric radio propagation and associated signal processing. Stress was placed on simulation of oblique propagation at HF by the most practical means, engineering the programs to achieve economical operation. The computer processes consist mainly of ionospheric structural analysis and coding, raytracing, and the subsequent use of ray trajectory data to simulate the operation of complete ionospheric radio systems.

This report summarizes the conclusions which have been drawn from the work, including much technical data, some programming details, and hopefully some insight into the optimum methods for doing this work. Also, there is a considerable amount of information inherently contained in these data which bear on the understanding of the propagation simulated. Many applications of raytracing are discussed, and many useful programming or plotting techniques are described, some for the first time. Finally, a number of plotted examples are given to show what can be achieved by these methods and to illustrate selected aspects of ionospheric propagation.

CONTENTS

ILLUSTRATIONS

ILLUSTRATIONS (Cont)

ILLUSTRATIONS (Cont)

TABLES

ACKNOWLEDGMENT

The work reported here was supported by the Advanced Research Projects Agency through the Office of Naval Research, Contract Nonr-225(64).

I. INTRODUCTION

It is fortunate that HF radio waves propagate through the ionosphere in a manner which is largely in accordance with the laws of geometric optics. As a consequence, we can study such propagation through the use of "rays" which are fairly easy to calculate. Were it not for this, it would be necessary to calculate the progress of energy using the wave equations which are a great deal more difficult to use. Recently, Kelso [1] stated the situation as follows: "If one examines the basis of radio wave propagation in the ionosphere from a position of considerable generality, the foundation for raytracing seems to rest upon an almost unending sequence of approximations. Fortunately, however, it can be shown that for frequencies in the HF spectrum--these approximations are generally well founded."

Since 1963, we of the Radioscience Laboratory have been calculating rays by digital computer techniques and using them for a variety of research studies. Our emphasis has been primarily upon the applications of computed rays rather than upon the methods for finding the rays. There is one exception to this generality: We found it necessary to invent methods for calculating rays through the use of Snell's law because such methods offered attractive economies but were not being exploited elsewhere. For the computation of rays by other approaches, we have chosen to use adaptations of existing raytracing programs.

All raytracing methods rest on the same foundation which can be stated mathematically in a variety of equivalent ways. For example, Fermat's principle alone could be used to derive all other equations used for raytracing. However, a variety of mathematical approaches are used because each has logical or computational advantages. Nevertheless, it must be stressed that all are based on the same, simple underlying principle. This underlying unity is difficult to detect in the details of the existing raytracing techniques because the superficial variations are extreme. However, every technique should lead to the same result regardless of the details involved in the calculation method.

To calculate rays, it is necessary to make a number of compromises based on the considerations of accuracy, cost, versatility, simplicity, acceptable physical assumptions, acceptable mathematical assumptions, and so forth. Since individual judgments vary in the weighting of these factors, there exist today many varieties of raytracing programs developed by different individuals. The newcomer to raytracing often finds himself deluged by different programs, each accompanied by claims of superiority, and the result is frequently the generation of confusion rather than insight. It is hoped that this paper will give the reader some perspective into the whole field of raytracing so that he may intelligently select the best program for his needs, or if necessary, create a new one to represent a new set of compromises.

In the ideal situation, all types of programs should be available to each user so that each individual problem can be solved with a program best adapted to it. This is usually not practical because of the continuous change in computer hardware and software within a single organization and computer nonuniformity in different organizations. As a consequence of these factors, it is not a straightforward matter to acquire someone else's program in a usable form. Usually, each person must do a considerable amount of modifying and testing before a new program can be run with confidence in a new facility. The unfortunate result of this is that only the largest users of raytracing can afford to keep current versions of the many different programs in readiness.

II. A REVIEW OF OBLIQUE RAYTRACING*

A. Forms of Raytracing

An outstanding characteristic of raytracing is the diversity of its forms. A ray calculation may involve the manual evaluation of a single equation or it may involve millions of calculations by a modern digital computer. The geomagnetic field may or may not be included; the ionosphere may be described in one, two, or three dimensions; the ray may be calculated in two or three dimensions; the process can be carried out on a digital computer, an analog computer, or by graphical or analytical techniques. Many levels of complexity exist; the choice among them is determined by the state of knowledge of ionospheric structure and by the intended applications.

B. Scope of the Subject Considered Here

Because of the wide variety of material which could be included here, let us begin by defining (and thus limiting) the subject to be discussed. Raytracing is that process which determines the path of a ray in a medium possessing a specified, inhomogeneous refractive index. The index may or may not be isotropic. Associated with the geometric optics concept of the "ray" are assumptions about the spatial distribution of the refractive index and the distribution of the families of raypaths. The exact method of trajectory determination is not limited in any way by this definition.

Here, we will concentrate on radio rays with frequencies above the LUF but in the high frequency band. These frequencies lie roughly between 2 and 30 MHz and thus exceed the electron gyrofrequency. Also, we will consider only oblique propagation; consequently, the index of refraction will be a fairly well behaved parameter along the paths of radio

*This chapter is a modified version of "A Review of Oblique Raytracing and Its Application to the Calculation of Signal Strength," presented by the author at the 11th Symposium of the AGARD Electromagnetic Wave Propagation Committee, Leicester, England, Jul 1966.

waves of interest here, contrasted to the complexity encountered in vertical incidence work. With these restrictions we avoid many of the problems which arise at vertical incidence and at frequencies very near or below the electron gyrofrequency.

Figure 1 illustrates the general behavior of the refractive index for oblique rays in the 4 to 8 MHz range. Raypaths have been digitally computed and automatically plotted by using Snell's law and the common no-field approximation for the refractive index $\mu^2 = 1 - (80.6N/f^2)$. The ionospheric model was a Chapman layer with a critical frequency of 4 MHz, a height of 300 km, and a scale height of 100 km. On each plot lines are drawn at the levels of various refractive indices, the point of interest being that μ is usually much nearer 1 than 0. Note that we could restrict our attention to the range $0.5 \leq \mu \leq 1.0$ and still be able to include most of the oblique rays from the illustrated set. This conclusion would not be materially altered by a change in the ionospheric model. Thus one can be guided by the general rule that $\mu < 0.5$, only if the takeoff angle exceeds 60° relative to the horizontal. Such rays return to earth only at frequencies near or below the ionospheric critical frequency. While the effect of the geomagnetic field complicates the behavior of the index, the general rule still holds approximately true for the real part of the index when operating frequencies exceed the electron gyrofrequency, and this usually happens above the LUF.

C. The Raytracing Process

A striking recent trend in scientific research has been the widespread use of digital computers. Progress in this application has been so rapid that many workers now automatically associate raytracing with computers, and the graphical and "slider" techniques which have been common in the past might someday be considered to be outside the field designated as "raytracing." Computers have been brought to a level of sophistication sufficient to permit their routine use for solution of long, involved problems. Such processes had previously been impracticable. Because of their obvious utility, computers are rapidly becoming

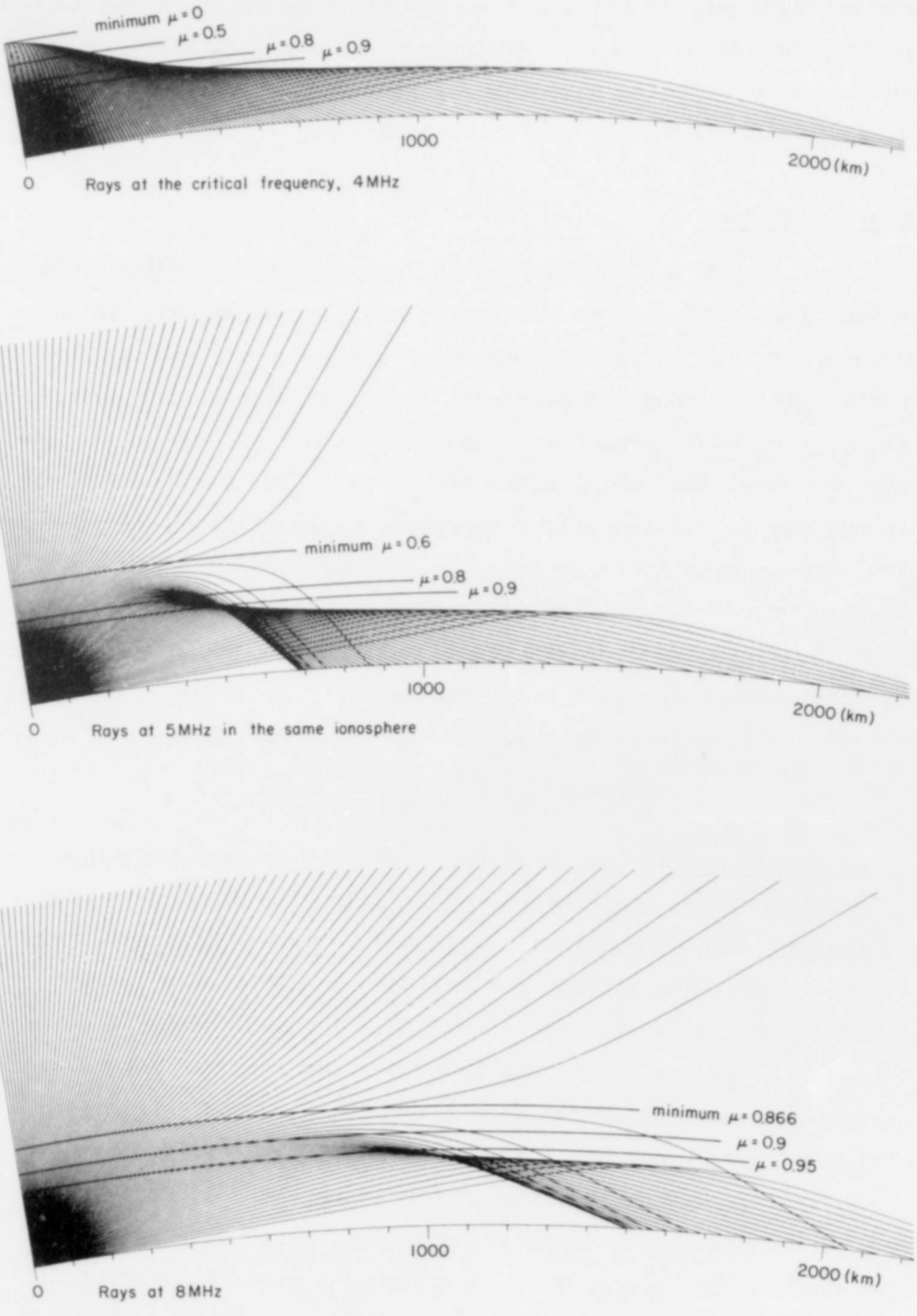


Fig. 1. RAYS AND REFRACTIVE INDEX LEVELS AT 4, 5, AND 8 MHz.

widespread so that many scientific research personnel have access to one. Those who work on the machines in depth soon realize that they are little more than very fast bookkeeping machines; thus computers are well suited to the execution of repetitive processes such a raytracing.

D. Computing Methods

Raytracing is similar to bookkeeping in that, after a propagation path has been initiated at some particular angle, the computer is instructed to accumulate in various "ledgers" the values of appropriate parameters as the ray progresses through the ionosphere to its destination. The most commonly computed ray parameters are the ground range traversed, the group time delay along the raypath, the number of wavelengths along the ray (often called the phase path), Faraday rotation, absorption, the geographic or geomagnetic coordinates and orientation of the raypath, etc. The primary difference between raytracing methods lies in the equations and approximations utilized at each step of determining the incremental parts of a ray trajectory. In carrying out this determination, three methods have gained widespread acceptance in recent years:

(1) Analytic Layers. The most rapid and inexpensive techniques make use of an ionospheric model consisting of two or three analytically described layers. Equations are derived which permit the calculation of a ray through all layers by the evaluation of very few expressions. In one program (known in the U.S. as the Kift-Fooks program after its originators at Slough), the ionosphere is modeled by three concentric parabolic layers, one each for the E, F_1, and F_2 regions [2,3,4]. A ray is computed from the ground through the lower layers to refract off the F_2 layer and return to earth; the entire calculation can be done by evaluating only seven expressions. On an IBM 7090, this method produced approximately 100 ray hops per second. This economy is offset by the severity of the approximations involved. In particular, the ionospheric

model is very crude; the magnetic field is neglected; the computation is only two-dimensional, and even one of these dimensions (the horizontal) can be accounted for by use of some very rough approximations. Attempts to overcome these deficiencies have produced some promising results [5], but there is still room for considerable improvement.

(2) Snell's Law. If the magnetic field can be neglected, the index of refraction can be treated as a scalar, and Snell's law may then be used in one of its trigonometric or differential forms. The ionosphere can be modeled by a very finely divided series of shells, or some other coding scheme may be arranged, so that a particular ray is followed through a two- or three-dimensional scalar field of refractive index by means of a large number of Snell's law solutions. There may be 200 to 1000 such calculations for a single hop of a single ray; nevertheless, each one is so quick that on an IBM 7090 one can compute from 1 to 25 ray hops per second. The author has been most active in this field because of an interest in the upper HF band where the exclusion of the magnetic field is often not a serious matter. However, when the complexity of the ionosphere or the complexity of the associated data processing becomes severe, or when the magnetic field is desired, the following method may be more appropriate.

(3) Haselgrove Equations. When the geomagnetic field is included, the refractive index becomes a function of both the field direction and the wave direction. One must then use more subtle mathematical techniques, the most popular of which has been that developed by the Haselgroves [6,7]. They developed six simultaneous differential equations which may be evaluated at short intervals along a ray to determine the successive ray direction changes. A more direct and straightforward approach is difficult because the index of refraction depends in a known manner on the wave direction but, during computation, only the

ray direction is known. If the index of refraction could be computed from the known ray direction, a faster computer process might be devised. There have been promising developments in this direction [8] but, to the author's knowledge, no practical computer processes have yet made use of such an approach.

It is possible (and apparently practical [9]) to use other equations for raytracing in anisotropic media. One can begin with Fermat's principle or with the equation of the ikonal [1] or with some other approach based on the methods of geometric optics [10,11]. While some particular derivation may yield more insight than others, each leads to equations suitable for use as a basis for computation, and the results should be identical. The choice among them is generally dictated by practical considerations of numerical analysis and programming, since the same fundamental mathematics is used for all. The most popular choice has been the method introduced by the Haselgroves.

The Haselgrove equations, combined with the appropriate refractive index equations and other information, can be used to compute a three-dimensional raypath in a three-dimensional electron distribution, taking into account a fairly realistic magnetic field. It is still difficult to incorporate these procedures in a computer program which is reasonably versatile and yet fast enough so that the cost is not excessive. There have been some very promising developments in this area. To a rough order of magnitude, the computation of a ray hop using the Haselgrove technique required from 2 to 100 sec on a 7090. (Computation times are given because of the recognized need for estimates, but they should be used with caution because times vary greatly depending on step lengths, data extraction processes, etc.)

This technique may be used without including the magnetic field, if the higher cost can be justified by considerations of expediency. (It may be cheaper to run this program than to develop

a Snell's law program, if the work load is light.) One of the disadvantages of this application is the need for well-behaved spatial derivatives of the electron density. If abrupt gradient changes occur in the ionosphere, the Haselgrove equations can be applied but this is difficult; a Snell's law program is comparatively unaffected.

The three raytracing methods listed above are widely used but other techniques can be used as well. There are several methods for raytracing with the inclusion of a geomagnetic field; two examples are the graphical method of Poeverlein [12] and the Booker quartic [13]. When the field is neglected, Snell's law can be applied in a variety of ways; in fact, the parabolic layer method of paragraph (1) above is, in a sense, a specialized Snell's law solution.

While the raytracing process is a mixture of mathematics and bookkeeping, its applications are usually of a practical nature. Each specific application determines which approach to programming should be taken. The author maintains in operational status several different programs, each of which represents an attempt to meet different needs by different means. When the sole information about the ionosphere must be taken from predictions, it is reasonable to use the Kift approach because its level of approximation is consistent with the use of such skimpy ionospheric electron distribution data. Alternatively, if a series of soundings are available along a particular path, one might use a Haselgrove program at 3 MHz or a Snell's law program at 20 MHz. If one needs only gross propagation information (such as might be needed for computation of sources of ground backscatter), then the Snell's law methods can provide sufficient accuracy at the least cost.

E. Non-Digital Methods

The current emphasis in raytracing by digital computer may be only a temporary situation. Special-purpose analog computers have been designed and built to raytrace, but their success has not been great and it is surmised that the approach lacks versatility. Such a machine might

be very economical, however, and it may come into use in the future. Work has been done on general-purpose analog computers [14], but here again the success has not been great. The prime barrier seems to be the difficulty and the inherent approximations of the programming needed for insertion of the electron density distribution. There is also a serious limitation of accuracy when analog methods are used. For rays to be used in the calculation of signal strength, errors become magnified when differences are calculated and so, for this application, analog raytracing has not been successful.

Some noncomputer methods have been used, notably the graphical constructions and the analytic solutions for simple ionospheres such as parabolic layers. Because of the amount of manual labor involved, the graphical technique is not practical when many rays must be followed through much ionospheric detail. However, the method does give excellent insight and some useful quantitative interrelationships when it is applied to a nearly homogeneous region.

The analytic techniques are most useful when the desired results, usually a relationship among various ray parameters, can be derived and evaluated without computing great numbers of rays. Also, as will be mentioned later, the analytic techniques can provide exact answers which may then be used as "primary standards" for checking the accuracy of computer-integrated ray calculations. Nevertheless, it does appear that for the present and near future, digital computers provide the best means for obtaining realistic raypath information.

F. Electron Density and the Choice among Methods

Newcomers to the subject of raytracing frequently associate it with an inherent ability to determine the electron density distribution at some real time and place. For example, a new user may say, "Please compute all the rays between London and New York at noon on July 16th." If the raytracing specialist replies, "You must first tell me the electron density distribution," he is likely to be regarded as some sort of fraud because he cannot perform his specialty!

There is a closely related and entirely valid question which should be carefully considered by anyone who intends to develop a raytracing capability. The question might be phrased as follows: "When I finish developing my program, the input data concerning electron distribution will suffer in some degree from a lack of information. In view of the expected quality of the input data, how much quality should be built into the computing technique?" Raypath computation is a means to an end, not an end in itself, and so the goals should be examined to determine the best method for reaching them. Excessively elaborate computation schemes bring undue expense and may convey a false impression of accuracy if the starting data are of low quality.

In examining this question, we should recognize that the problems to be solved fall into two classes:

(1) Some problems concern rays in a theoretical ionosphere such as a Chapman layer, and it is desired to have results which are reproducible by other methods or even by other workers so that appropriate comparisons can be made. In these problems, the upper limit of a desired accuracy is set by weighing the patience and resources of the operator against the likelihood that the data will be rendered invalid because of the presence of errors of some particular size. Specific results may be of lasting interest for this class of problems.

(2) In other problems, one needs to know the raypaths which actually existed (or will exist) in some real situation. In this case, one must somehow obtain an ionosphere description before the ray calculation can be made. For example, if predicted data are the only source then it would be unrealistic to use a three-dimensional program including the magnetic field to determine the ray distribution. Rather, one might use the Kift method or perhaps a Snell's law approach with a long integration step. This situation represents perhaps the extreme case in which crude raytracing is justified.

Many intermediate stages exist wherein intermediate accuracy levels are desirable. For example, when ionospheric sounder records are available for the time and place of interest, one might choose a Haselgrove or Snell's law program depending on whether the frequencies are in the lower or upper parts of the HF spectrum. If elaborate ionospheric sounding data are available, then elaborate treatments may be justified, depending on the use intended for the computed data.

No single answer can be given to the question posed; no single raytracing program can meet all needs. The matching of raytracing techniques to specific problems is essentially one of engineering compromise. Options are available only to the person who has a variety of programs designed for different situations available from which to choose.

G. Specific Applications of Raytracing

Raytracing is essentially a tool which permits us to study radio propagation in an ionosphere of arbitrary complexity, limited only by the assumptions inherent in geometrical optics and the practical limitations associated with a particular technique. Applications usually involve some form of simulation and, in a sense, raytracing provides us with a means for carrying out controlled "laboratory" experiments. This is a particularly valuable asset because the real ionosphere is not under control, while a simulated ionosphere may be varied at will.

Sometimes ionospheric radio measurements exist but there is no information available concerning the ionospheric structure. In this case, a variety of hypothetical structures can be used as a basis for raytracing so that (hopefully) by an iterative process, one can reproduce the measured data and thus explain its meanings. Often ionospheric structural information is available in some form and one wishes to know where the radio energy travels, how long it takes, how much Faraday rotation there is, etc. In this case, all the ionospheric data available is used to make one ionospheric model which is then run through an appropriate raytracing program to produce the desired information. A few examples will be provided to give substance to these generalizations.

H. Determination of Individual Raypaths

Perhaps the most frequent application of raytracing is simply the determination of complete families of raypaths. Many situations arise in which one has ionospheric information and wishes to know where all rays go. A convenient approach providing this information calls for the initiation of a large number of raypaths which leave the transmitter (or receiver) at equally spaced takeoff angles. The relative spacing of

such rays provides a measure of signal strength. The computed information about raypaths is generally presented in one of three forms, depending on the means of computation and the needs of the user.

1. Numerical Descriptions of Raypaths

If the computation is analytic, graphical, or by digital computer, it is possible to express the ray trajectories in a written series of numbers by some code. While this is often the easiest form to produce, it may be the least useful. Digital computers can produce pages of raypath information, but it is difficult for a reader to obtain much insight from tables of numbers. Nevertheless, if only a few rays are involved, the written format may be the most convenient means for gaining an objective.

2. Graphically Displayed Raypaths

If a ray is computed in two dimensions and the user of the information does not have to carry out subsequent numerical computations with raypath results, then perhaps the most convenient format for raypath display is graphical. Figures 2 and 3 show many rays computed digitally and drawn by an automatic plotter. This presentation method conveys immediate insight to the reader. If the rays are computed by analog techniques, the plotted data are usually the easiest to produce.

When the raypath information is to be used in some calculation where extreme precision is not required, then the plotted information may be used as part of a graphical computation scheme. Such graphical methods are inherently rapid, and (perhaps more important) it is less likely that large errors or inconsistent logic will be used in a graphical calculation than in the same calculation performed by manual operation with numbers.

3. Raypath Information Coded as Computer Data

Situations often arise in which digitally computed rays are to be processed further by computer. In this case, the raypaths can be stored by some method so that they can be reinserted in the computer at a later time. It is the author's impression that he is the pioneer of

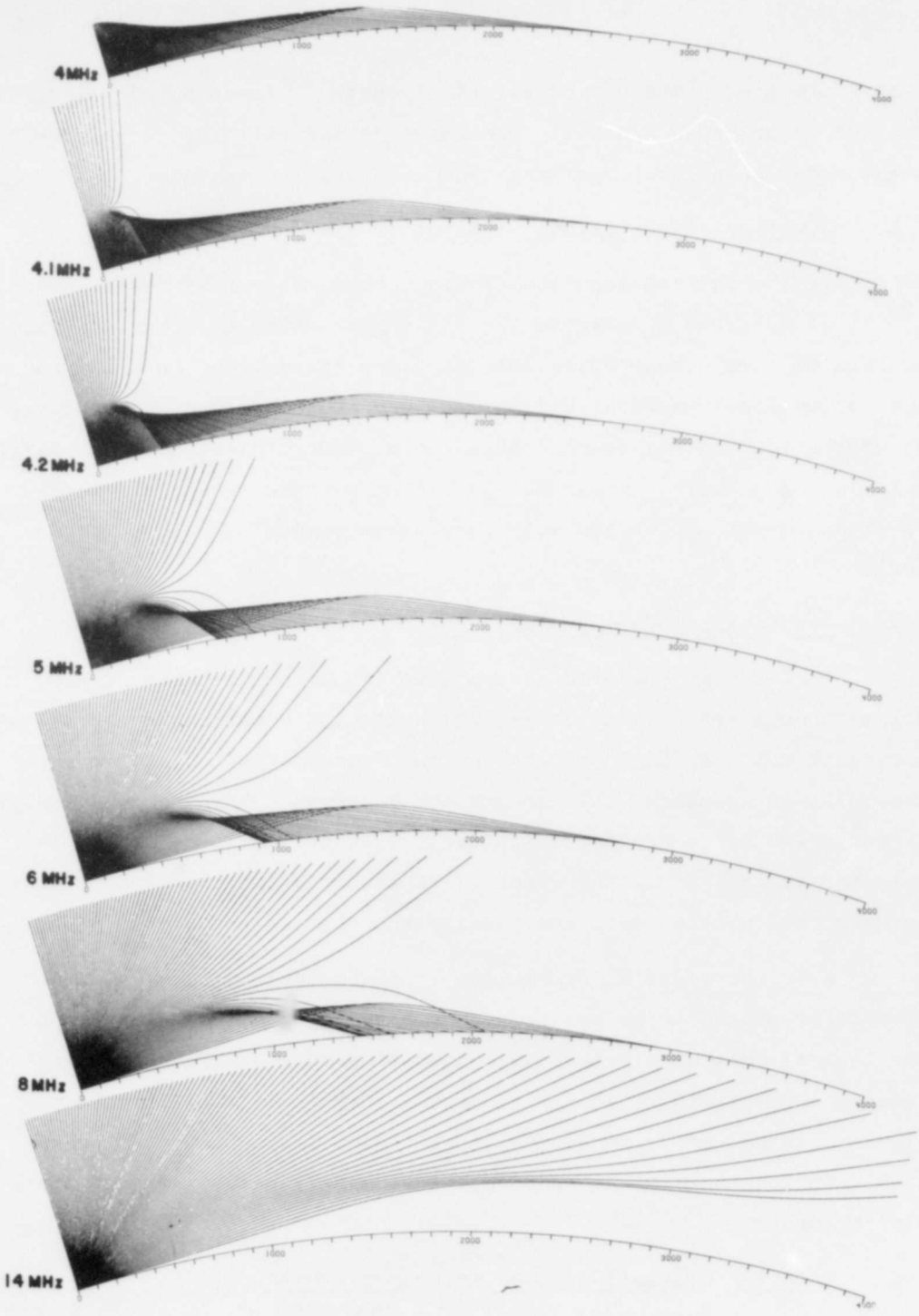


Fig. 2. EXAMPLES OF CALCULATED RAYPATHS IN A CHAPMAN LAYER.

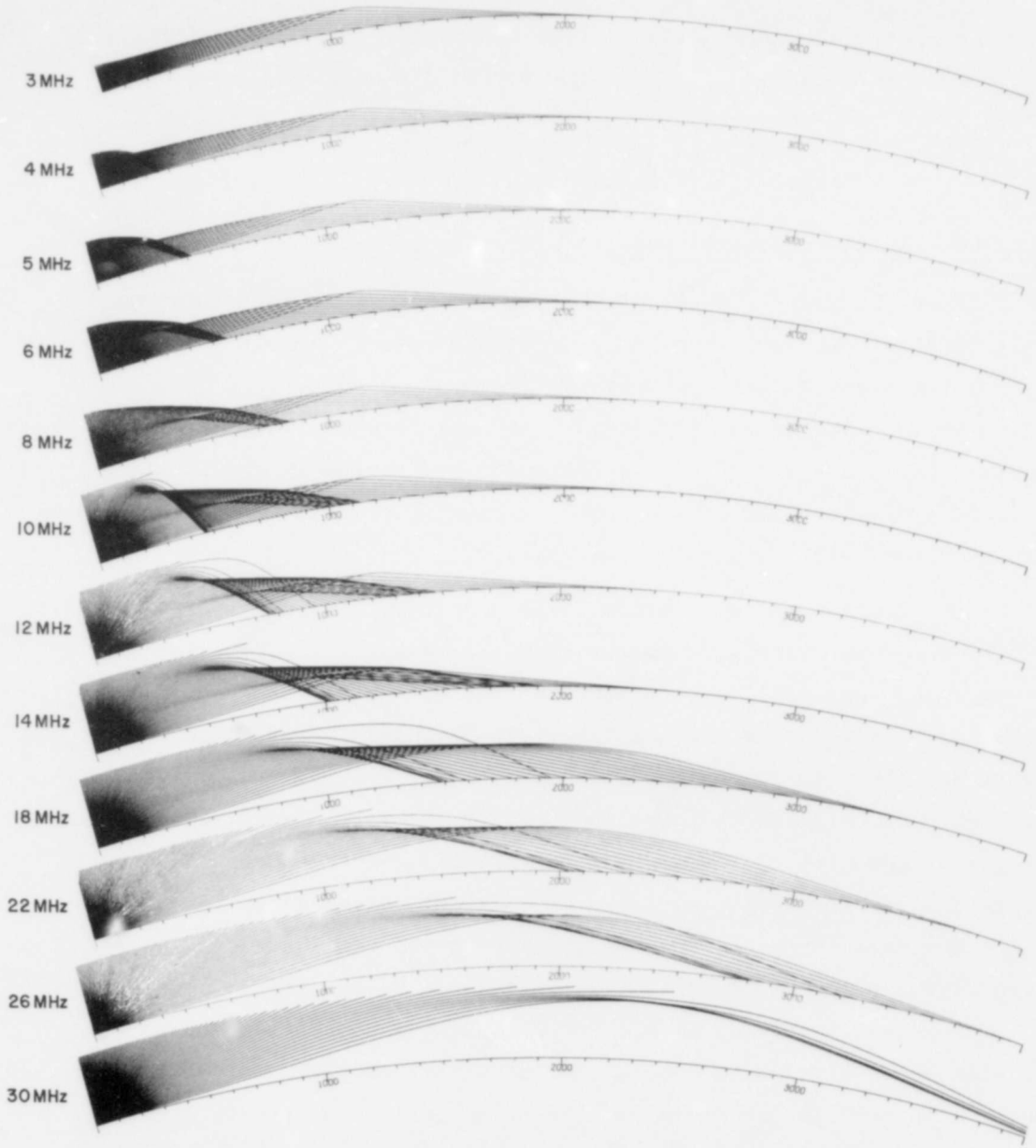


Fig. 3. RAYPATHS IN A TWO-LAYER IONOSPHERE.

this technique, whereby ray trajectories are stored on cards (called "raysets") which are used as data for subsequent programs. The coding scheme, which involves one punched card per hop per ray, permits large families of computed rays to be used over and over as data for new and

different computer analyses. The storage technique rapidly pays for itself because the computed data, easily stored and retrieved, can be reused as often as needed. Figure 4 illustrates the author's rayset coding method.

I. Computation of Transit Time for Analysis or Instruction

During ray trajectory calculation, particularly when it is carried out in a digital computer, it is easy to simultaneously compute transit time along the raypath. For radio waves in the ionosphere, two different transit times can be defined through the use of phase velocity and group velocity. They are called the "phase delay" and the "group delay"; often these are multiplied by the speed of light (299.7925 km/msec) to obtain the "phase path" and the "group path."

Like the trajectory information described above, the time delay data can be presented in written, computer coded, or graphical format. The first two types are easily visualized, but the merits of the graphical presentation of time delay are not widely appreciated. Four examples of this work are shown to demonstrate their utility. Figures 5 and 6 show raypath families, similar to those in Figs. 1 through 3, in which additional lines approximately orthogonal to the raypaths are added. In two cases, these lines represent surfaces of constant phase delay. Such surfaces are separated from one another by equal numbers of wavelengths, and they are useful primarily in the calculation of doppler shift. For example, if a satellite passes overhead, transmitting some frequency, the signal will be received by a ground station with a doppler shift. The doppler can be readily predicted if this technique is used; rays starting at the receiver are computed to find the surfaces of constant phase delay in the vicinity of the satellite. Then, every time the satellite passes through a surface, the phase path to the satellite decreases by the number of wavelengths which separate adjacent surfaces. By this technique, the doppler shift can be computed without making any severe assumptions about the nature of the satellite trajectory or the ionospheric structure; the computation technique is probably more accurate than the ionospheric structural information usually available.

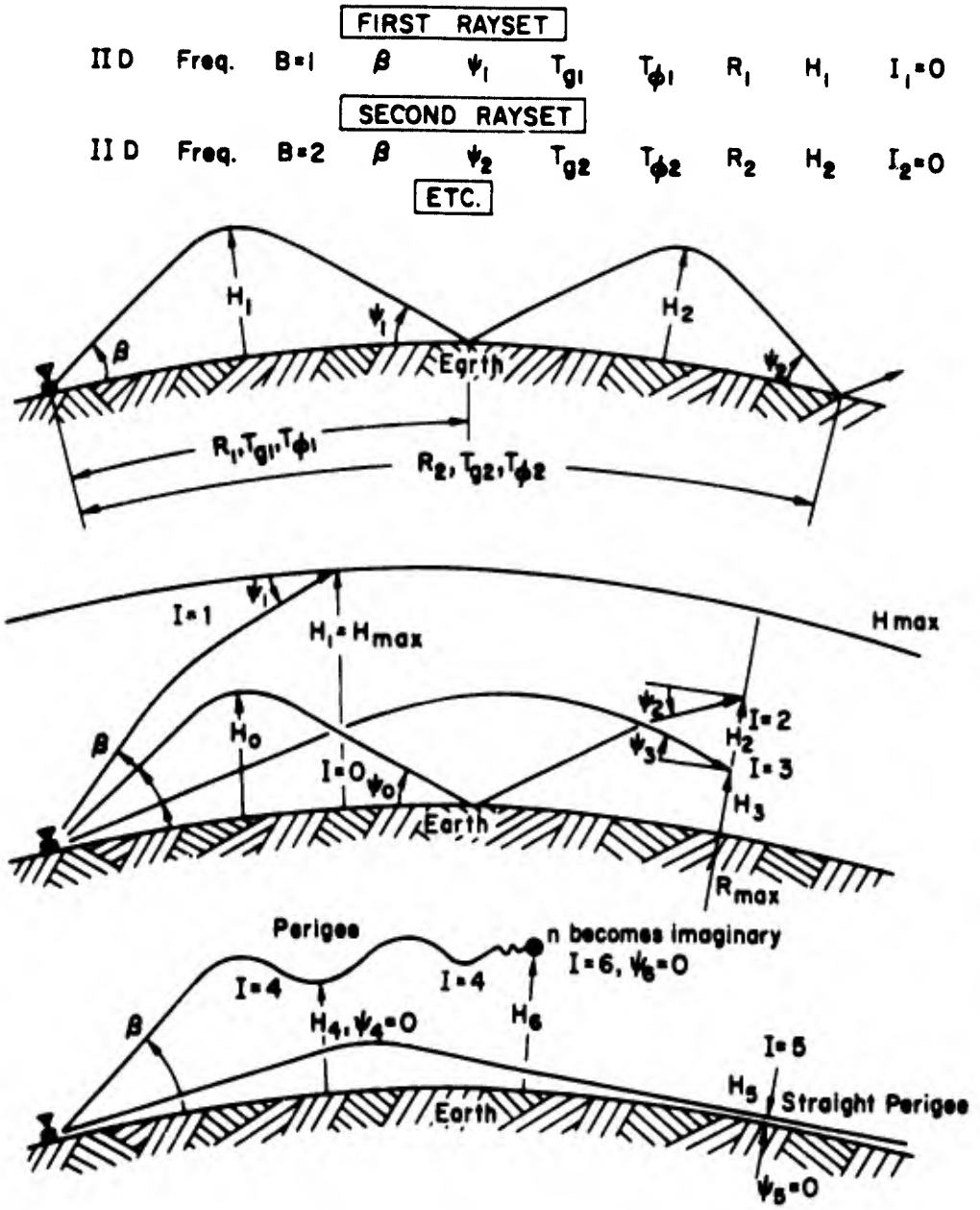


Fig. 4. THE RAYSET CODE, STORING RAYS FOR LATER DIGITAL COMPUTATION.

The other two ray plots in Figs. 5 and 6 show group fronts, which are lines of constant group delay. These are not necessarily orthogonal to the raypaths and, in fact, some are almost parallel to raypaths for short distances. The group fronts represent the successive locations of a short pulse of radio energy, or it might be said that they represent the transit time for information.

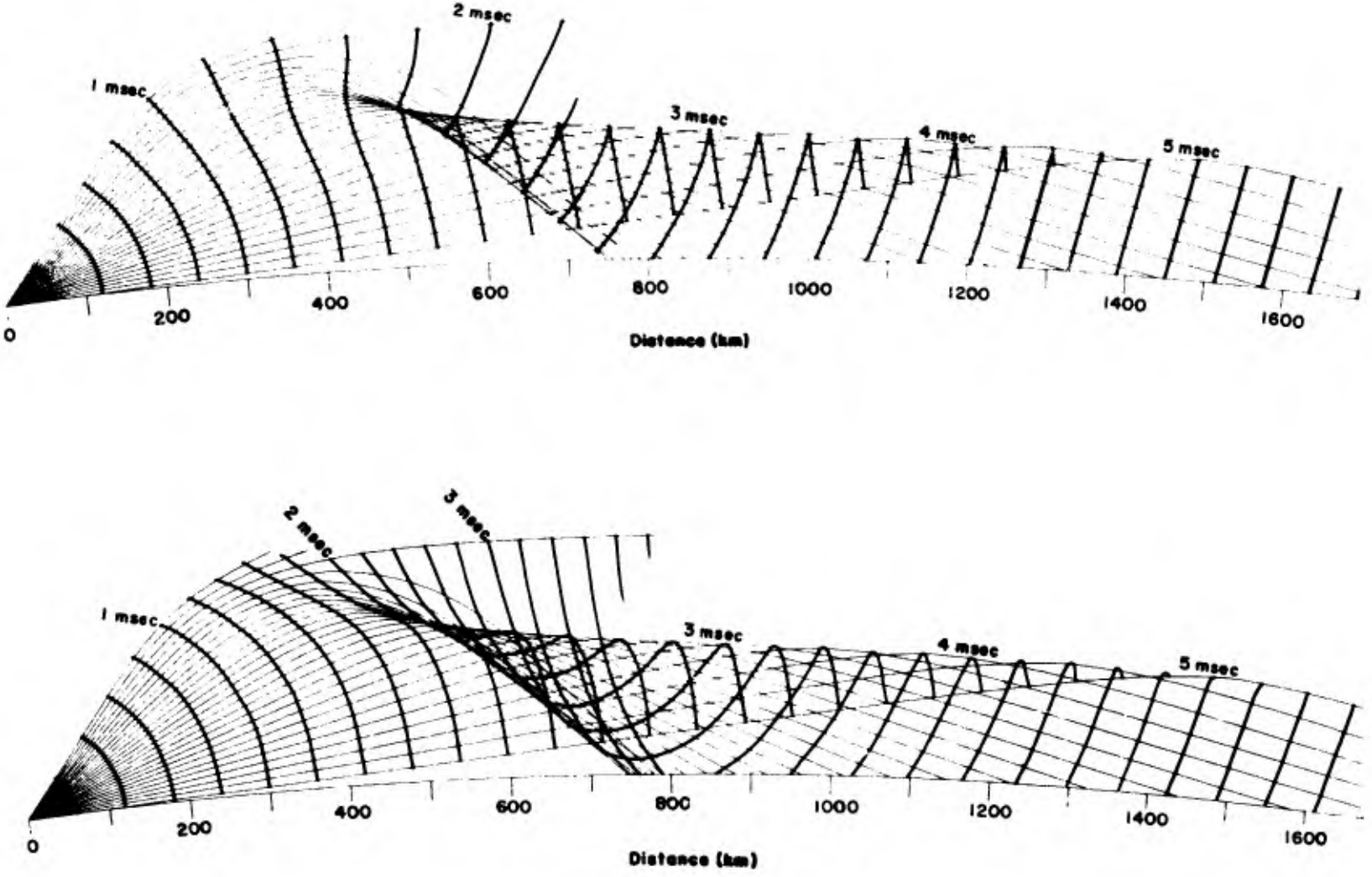


Fig. 5. SURFACES OF CONSTANT PHASE AND GROUP DELAY; SINGLE-LAYER IONOSPHERE.

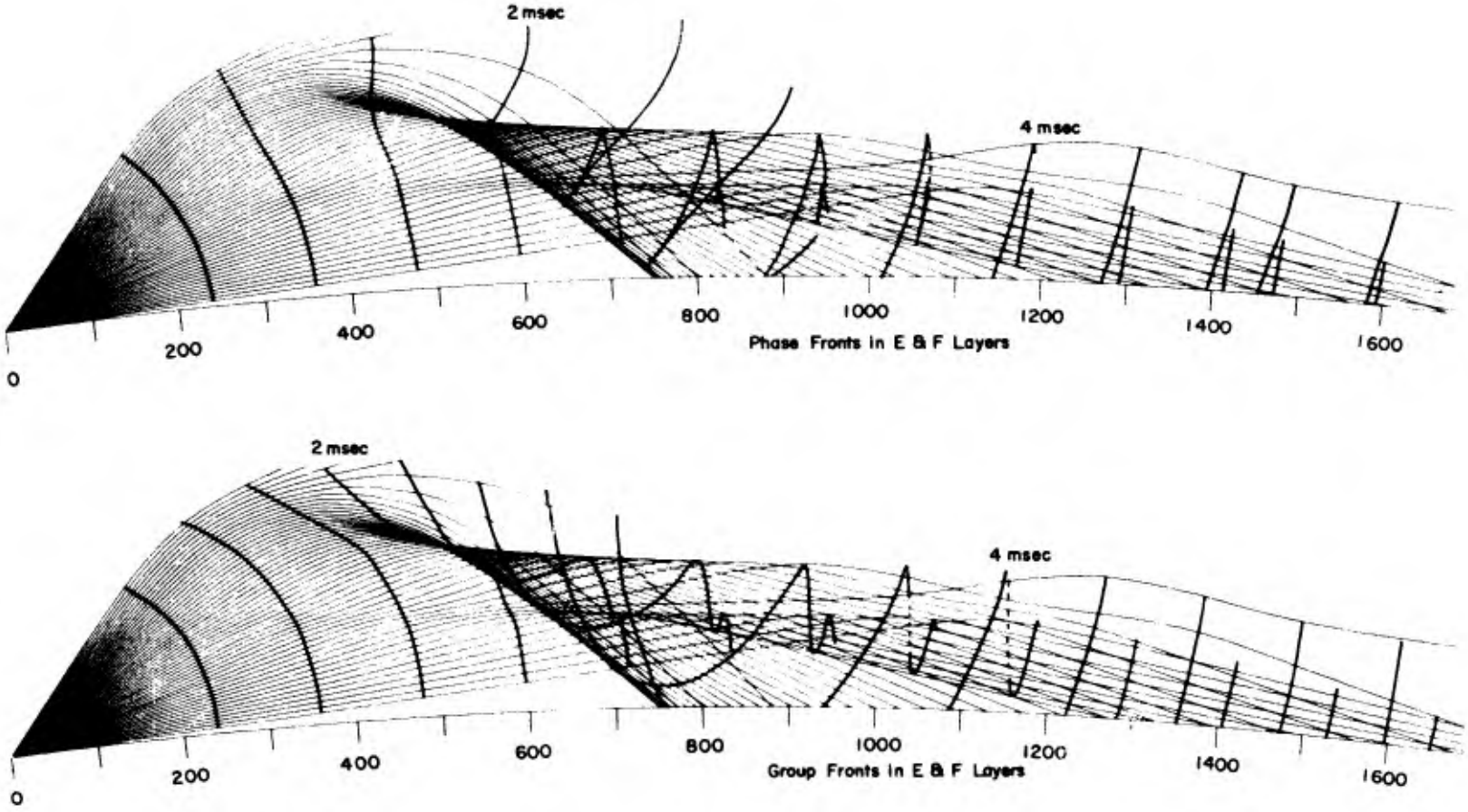


Fig. 6. SURFACES OF CONSTANT PHASE AND GROUP DELAY; DOUBLE-LAYER IONOSPHERE.

Because group velocity is defined by consideration of the sum of two continuous waves of almost identical frequency, the delay lines must be interpreted with some caution. When a short pulse is actually transmitted, it contains frequencies which differ appreciably from the carrier and, because of ionospheric dispersion, the pulse duration increases as it travels through the ionosphere. This dispersion is not shown by a single delay plot.

If a relatively long pulse of perhaps 200 μsec is transmitted and if the separation between adjacent group delay lines is also 200 μsec, then it would be reasonable to consider that the pulse at some later moment would be confined entirely between two adjacent lines on the plot. In this manner, one can visualize the travel of a pulse through the ionosphere.

Examine a matched pair of phase and group delay plots. Notice that near the origin the energy initially expands in circular (spherical) fronts. As the radio wave enters the ionosphere, however, the phase surfaces advance because of the increasing phase velocity. This advance is actually the cause of refraction, and such a plot provides insight into the reason for the action of the ionosphere on radio waves. Conversely, group fronts become retarded as the energy enters the ionosphere. The comparison between the two plots provides a concrete appreciation for the difference between group and phase time delays.

J. Faraday Rotation Calculation

Because of the different wavelengths and characteristic polarizations of the ordinary and extraordinary components, a linearly polarized radio wave entering the ionsophere is usually subject to a polarization change known as Faraday rotation. Figure 7 illustrates an example of this action. This rotation can be simulated by raytracing techniques, at least when digital computers are used. Essentially, there are two methods for proceeding. In the more rigorous but much more expensive method, both ordinary and extraordinary rays are computed to join the same two points on the earth's surface. During this computation, the

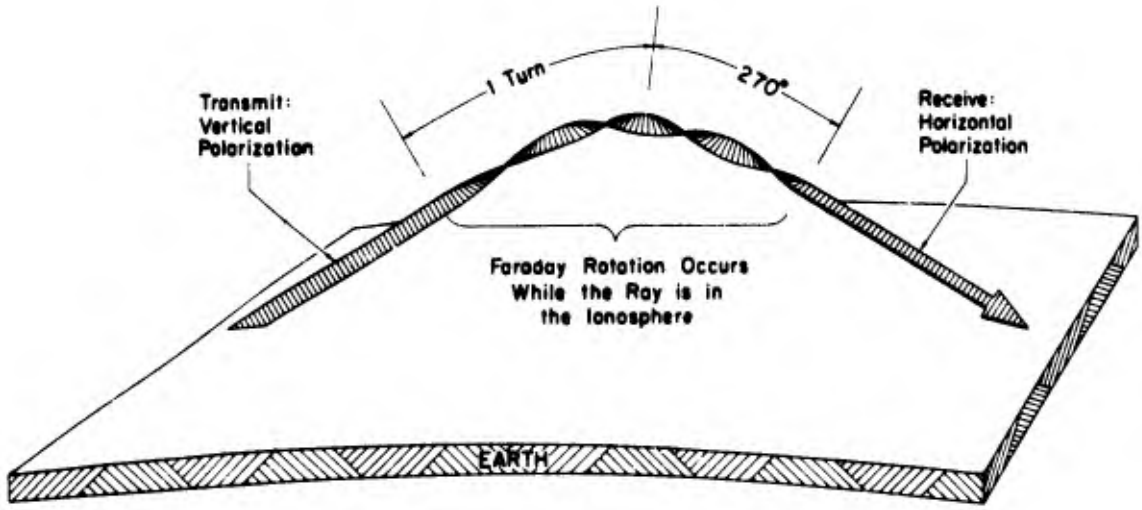


Fig. 7. ILLUSTRATION OF A POSSIBLE EXAMPLE OF FARADAY ROTATION OF THE PLANE OF POLARIZATION.

total number of wavelengths is accumulated for both modes. The difference between the two totals is just twice the number of Faraday rotations. This method is difficult because of the number of rays which must be computed to find the matched pairs which reach the same point. Accuracy requirements are stringent because the desired result is the difference between two large but nearly equal numbers. Thus, small percentage errors in the large numbers (the phase delays) produce large percentage errors in the small resulting difference. In addition, computations including the magnetic field cost roughly 10 to 100 times more than computations without it.

Fortunately, the amount of rotation can be roughly evaluated in many circumstances through the use of a Snell's law approach with a quasi-longitudinal expression for Faraday rotation. This approximation is commonly used in the study of the rotation of polarized VHF waves from passing satellites [15]. The main limitation of the technique for HF lies in the inherent assumption that the O and X rays stay close to one another. Computer results are accurate within about 5 percent if the ray takeoff angle is 30° or less. If the ionospheric model contains sporadic E (E_s), then the raypaths lie close to one another even when the takeoff angle becomes

steep; thus the approximation is useful up to about 60° in this case. Nevertheless, this more economical method must be used with caution and preferably with occasional spot checks by the first method.

Some results, computed by the second method, are shown in Figs. 8 and 9. The ionospheric models are shown in Fig. 10. Both models have an E_s layer and, in each, rays have been computed for three cases (geomagnetic latitude = 30°N, azimuth = 030°):

(1) Some rays pass through the E_s as if it weren't there; these refract through the "normal" ionosphere. The Faraday rotation

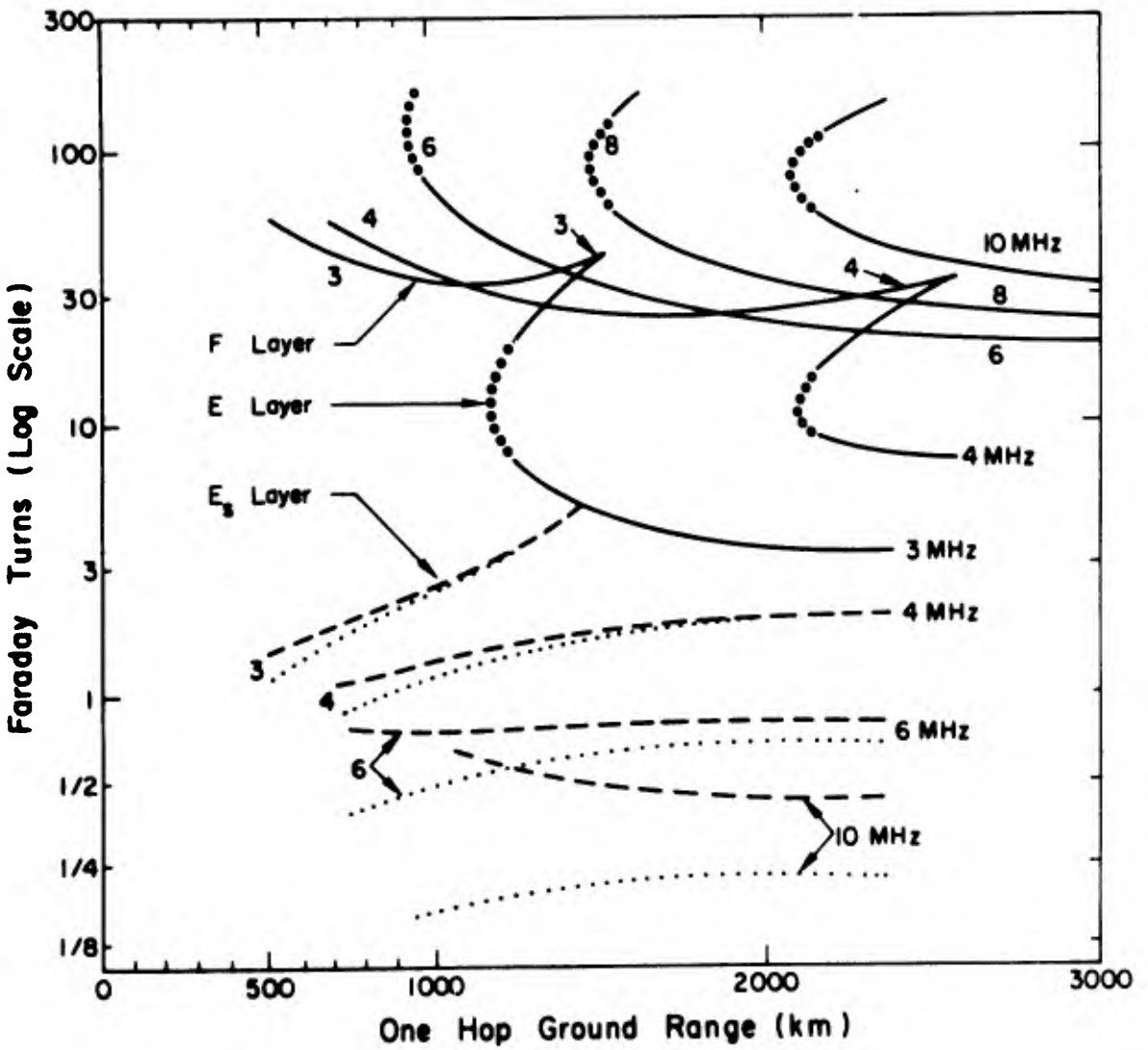


Fig. 8. FARADAY ROTATION AT NIGHT; E_s, E, and F LAYERS.

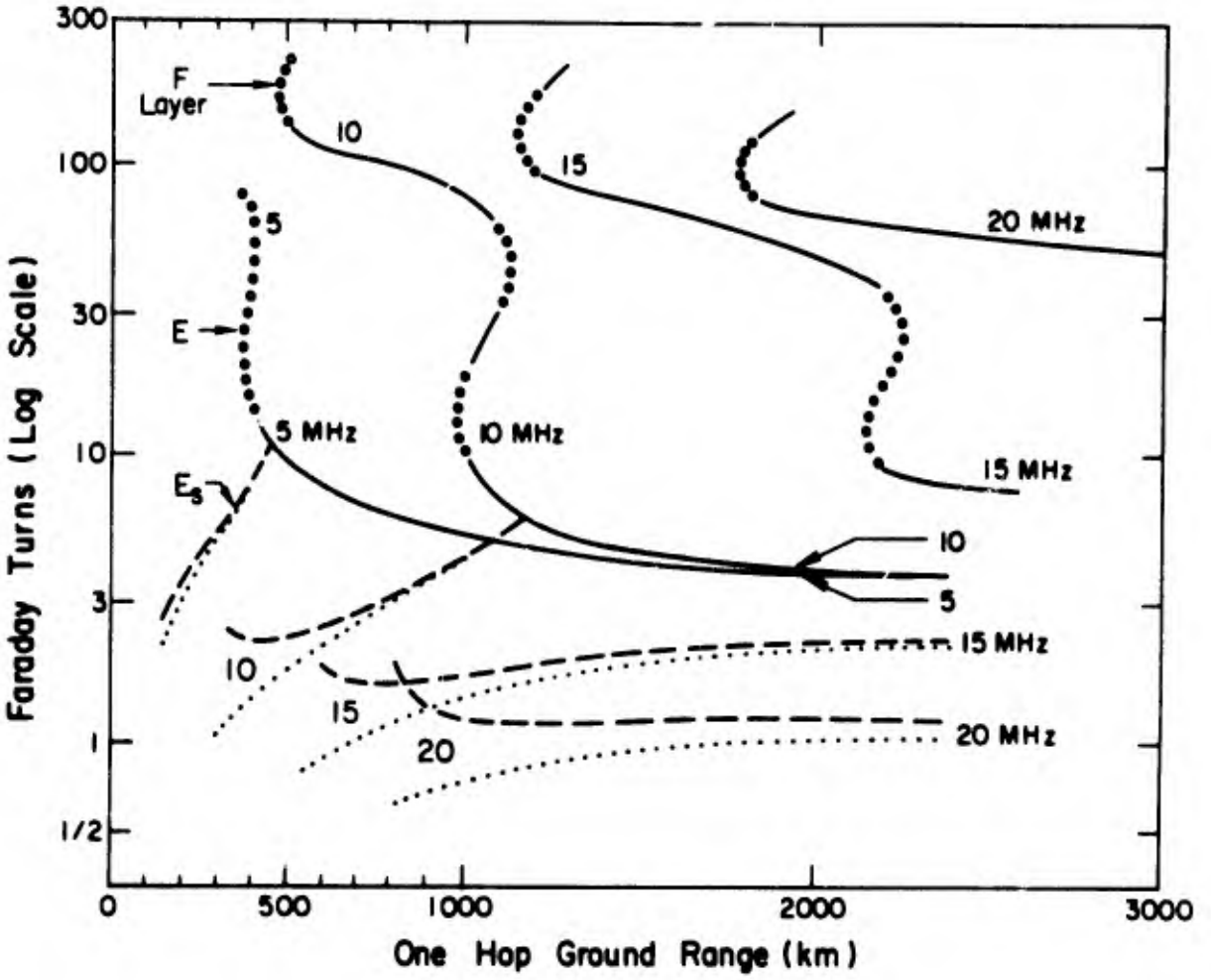


Fig. 9. FARADAY ROTATION IN THE DAYTIME; E_s, E, AND F LAYERS.

of these paths is plotted as a function of the range of the rays and drawn as a solid line for each radio frequency. When these lines become vertical, they are usually invalid because the unchanging range means that the skip distance has been reached. There, the difference between the trajectories of the O and X rays is so great that the first computational method must be used. The invalid portions of the curves are shown as a series of open circles.

(2) Other rays enter the E_s layer and refract by the Snell's law mechanism so that they return to earth. In this case, there is some Faraday rotation during the refraction inside the layer. These results are shown as dashed lines on the plots.

(3) A third type of raypath reflects off the bottom of the E_s layer because of the rapid change in the refractive index. It is assumed in this case that the index changes appreciably in a few wavelengths so that the geometric optics approximation

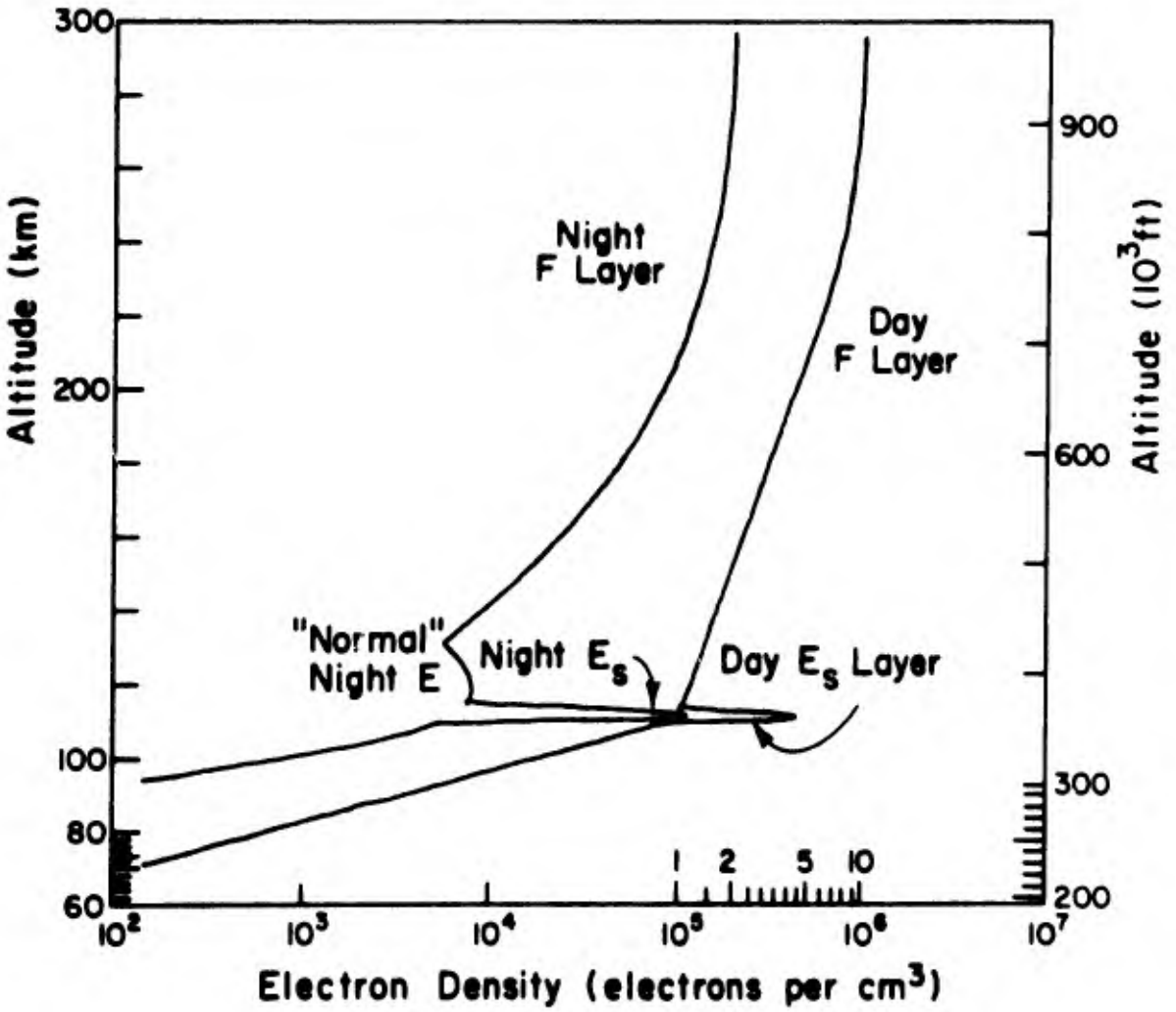


Fig. 10. IONOSPHERIC MODELS USED FOR COMPUTING FARADAY ROTATION.

breaks down. The wave solution would show an impedance mismatch with consequent reflection. This is simulated in the ray model by postulating a mirror at the base of the E_s layer, and the results are shown by dotted lines on the plots.

It is interesting to note that by this technique raytracing can be used even in the presence of refractive index structures which violate the assumptions inherent in ray theory. Such techniques can be used only if the full wave solution is known in advance and can be programmed into the computer. Of course, in the case of this E_s application, the preprogrammed solution is almost trivial. The same approach could be used, nevertheless, in more complicated situations (such as in the study of reflections from field-aligned structures) which violate the ray assumption.

K. The Synthesis of Oblique Ionograms

Oblique ionograms are records of sweep-frequency ionospheric sounders with transmitters separated from receivers by hundreds or thousands of kilometers. These ionograms cannot be inverted to give ionospheric structure without making rather severe assumptions concerning the nature of the ionosphere. However, ionograms can be calculated from a known ionsopheric structure in a straightforward manner through the use of raytracing. In essence, one calculates all the propagation modes which connect a hypothetical transmitter to a remote receiver and then plots the computed group delay vs frequency.

Because of the procedure used in most raytracing programs, it is usually not possible to specify the ground range at which one desires to land a ray. Rather, one must specify the orientation of the raypath as it leaves the transmitter and then calculate to see where the ray goes. After doing this a large number of times, it is possible (through interpolation techniques) to make a fairly good guess at the necessary starting ray orientation which will place the end of the ray at the desired spot.

The author's experience in attempting to implement this process has led to the conclusion that it is more practical to compute a complete family of rays at equally spaced takeoff angles. Then one can use interpolation techniques to find the final answer, rather than using the interpolation to find the starting orientation for calculation of a new ray. This method has worked well for ionogram synthesis and should work well for other applications which involve only raypaths which join two fixed positions on the earth's surface. The entire process of ionogram synthesis can be carried out in the computer, and synthetic ionograms can be drawn by a plotting machine controlled by computer. Two results are shown in the lower left of Figs. 11 and 12. Approximately 135 ionograms have been computed [16].

The author chose the Snell's law raytracing approach because the ordinary and extraordinary traces on an oblique ionogram are so similar to each other in structure. Computations without the geomagnetic field produce a single-component ionogram which exhibits the same structural

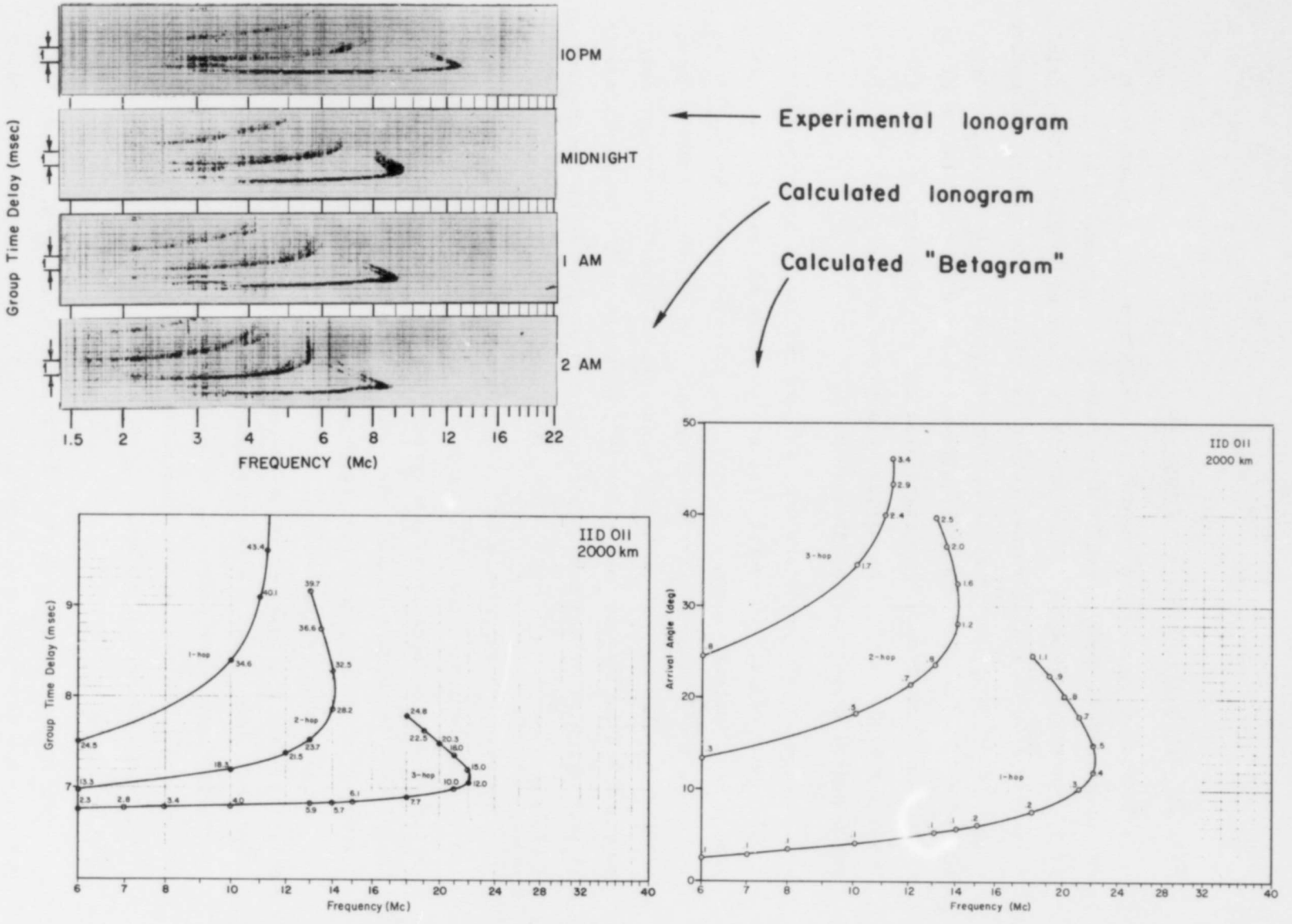


Fig. 11. EXPERIMENTAL AND SYNTHETIC SOUNDINGS; SINGLE-LAYER IONOSPHERE.

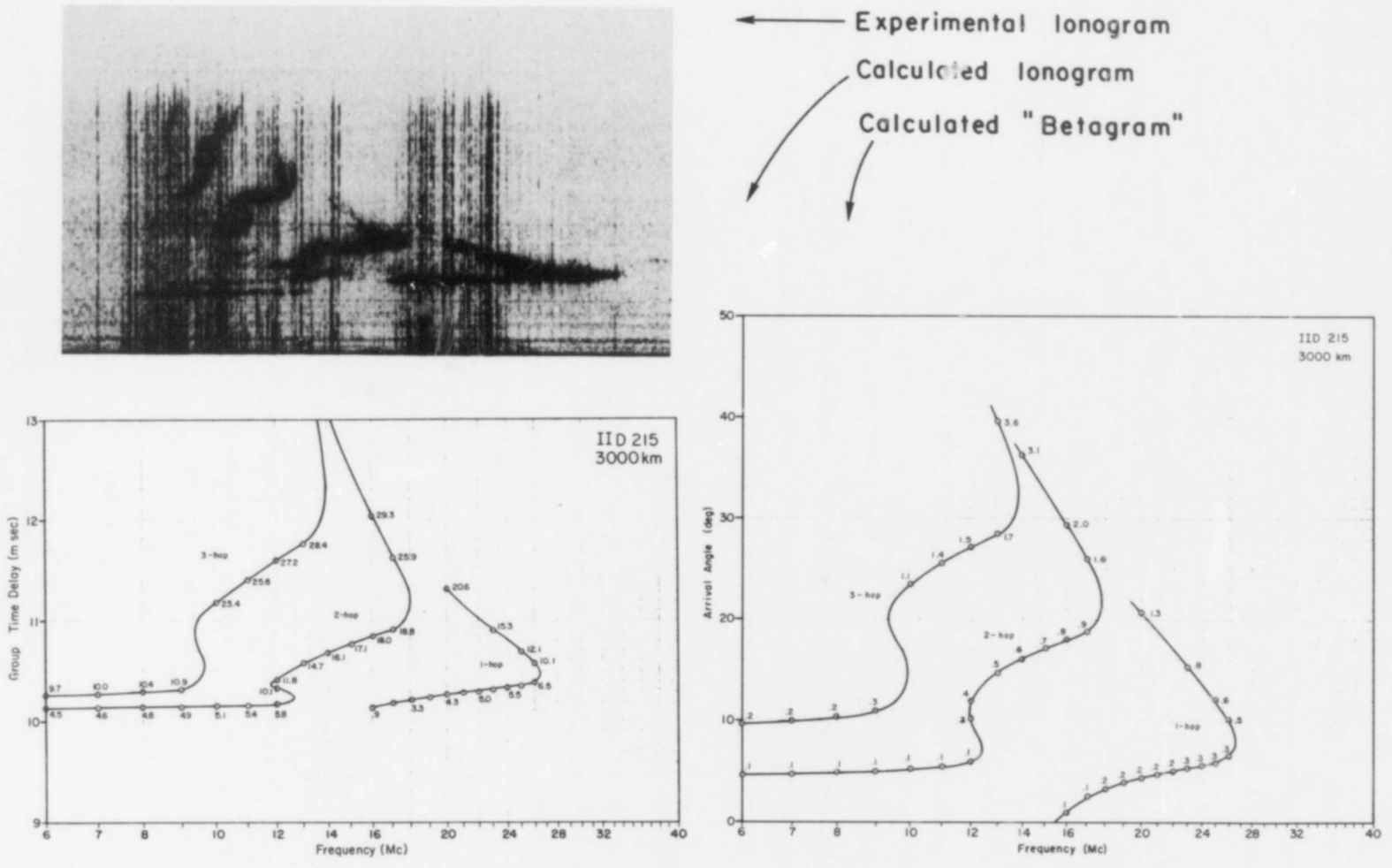


Fig. 12. EXPERIMENTAL AND SYNTHETIC SOUNDINGS; DOUBLE-LAYER IONOSPHERE.

characteristics as the O or X component on a real ionogram. This approach leads to high speed and economy, but it does not reproduce the small-scale differences between the O and X components which may be of interest in some applications.

In the author's program, signal strength has not been calculated, but this could be included by straightforward methods if desirable. The small numbers written along the synthetic ionogram traces are raypath takeoff angles measured relative to horizontal at the transmitter end of the path. One advantage of the synthetic process over the real counterpart is that the absolute magnitude of the group delay is revealed.

The major advantage of digital raytracing over other techniques for ionogram synthesis lies in the study of ionospheres containing lateral electron density gradients or localized irregularities. Analytical and graphical methods have inherent limitations in this application, but for digital raytracing the inclusion of such tilts and irregularities involves only the addition of more "bookkeeping."

L. Signal Strength of Sweep-Frequency Ground Backscatter

Figure 13 shows sweep-frequency backscatter records acquired by a specially calibrated pulse oblique backscatter sounder located at Stanford University [17,18]. These records contain much structural information which has been difficult to understand. Because of the variety of the backscatter structures encountered on these records, it is the author's belief that they must contain much useful information about the ionosphere, if only we knew how to understand the data.

1. Backscatter Analysis

Consider the backscatter propagation phenomenon. Initially, after a pulse is transmitted, it spreads in a hemispherical shape (Figs. 5 and 6). Eventually, as the energy travels up into the ionosphere, the hemispheric structure distorts and folds upon itself until the energy distribution in space and time is complex. When the energy reaches the earth at the end of the hop, it is scattered by a large number of ground reflectors. Each echoing surface initiates another spreading pulse of radio energy which again distorts as it travels back toward

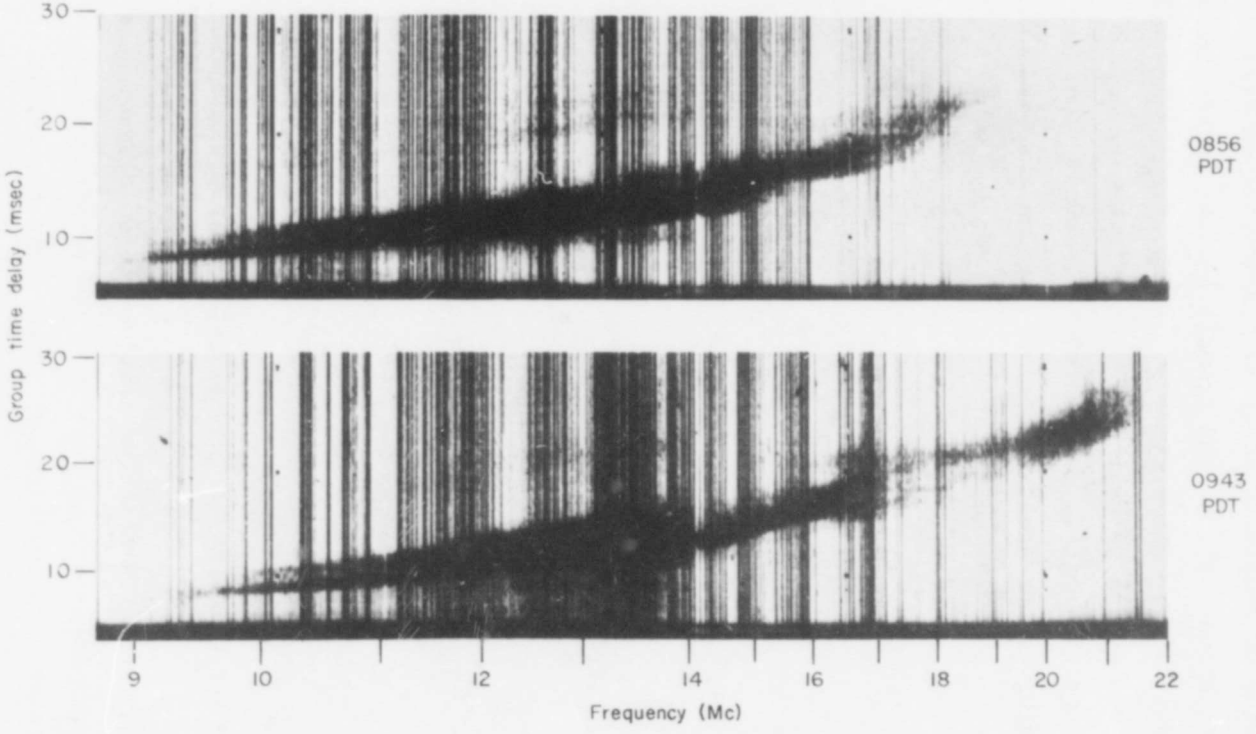


Fig. 13. TWO EXAMPLES OF SWEEP-FREQUENCY GROUND BACKSCATTER SOUNDINGS.

the sounder receiver, where it will eventually be detected as one component of the ground backscatter.

The simulation of this process is a complex application of raytracing to signal strength computation. The derivation is given in Ref. 17 following a line of reasoning which is illustrated in Figs. 14 through 18. Figure 14 shows the rays and symbols for the energy transit from the sounder transmitter to the scatterers on the ground. The energy is assumed to remain between raypaths 1_t and 2_t to strike the ground in area A_1 whose dimensions are readily calculated. Figure 15 shows the scattered rays returning to the sounder receiver. Since the simulation must make use of only precomputed rays which start from the sounder, it follows that raysets are not available for paths 1_r or 2_r which do not reach the sounder, but bracket it.

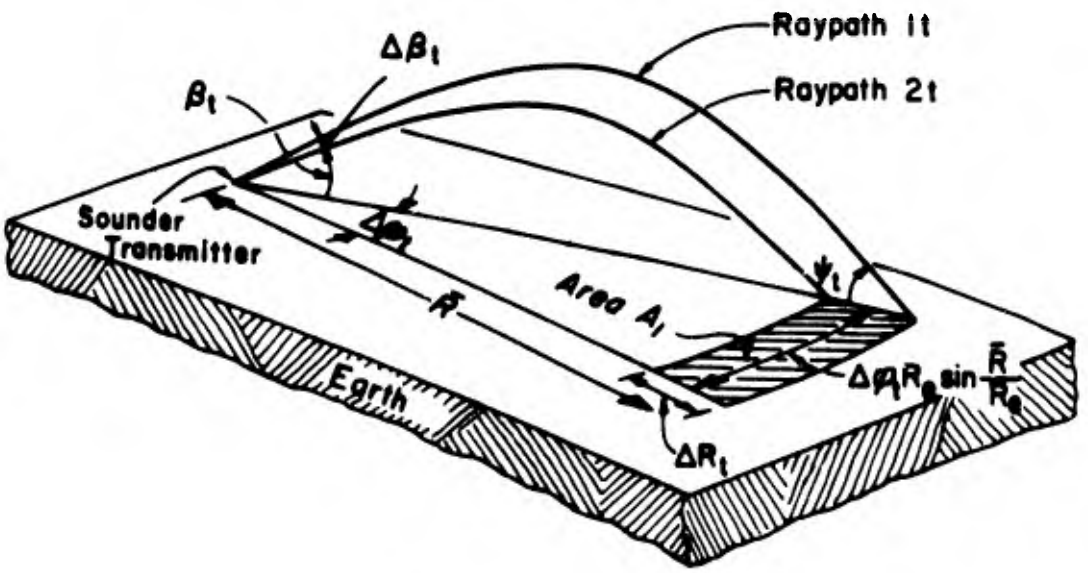


Fig. 14. METHOD FOR CALCULATING TRANSMITTED SIGNAL STRENGTH.

The signal strength calculation involves three dimensions: range, time, and power. These are used as the axes of a three-dimensional study, illustrated first by Fig. 16, which shows the hypothetical space-time distribution of energy after one-hop propagation through a single-layer ionosphere. Notice that on these graphs, volume corresponds to energy per meter, the meter being measured in azimuth.

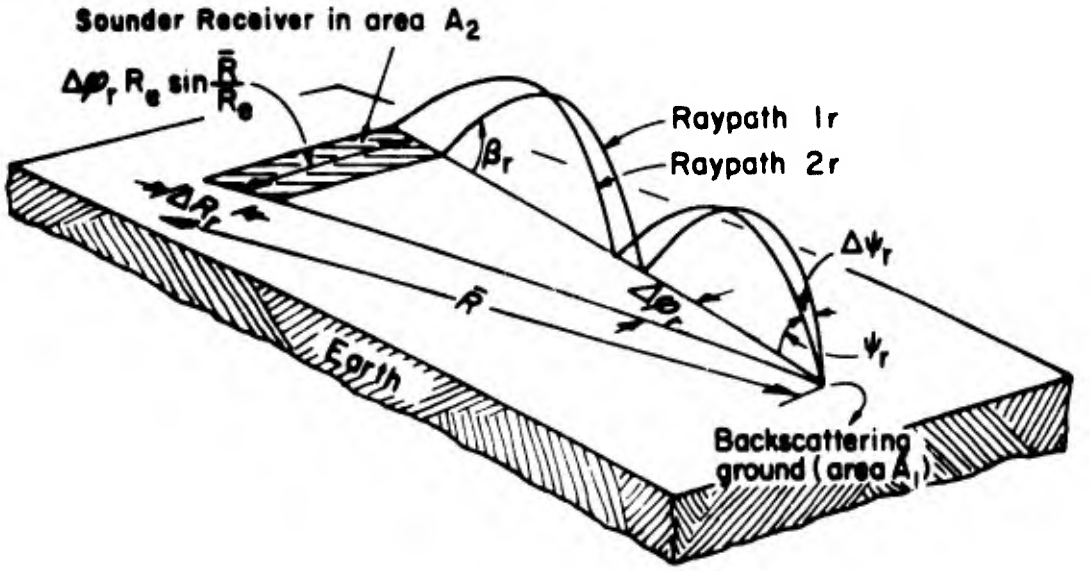


Fig. 15. METHOD FOR CALCULATING RETURNING BACKSCATTERED SIGNAL STRENGTH.

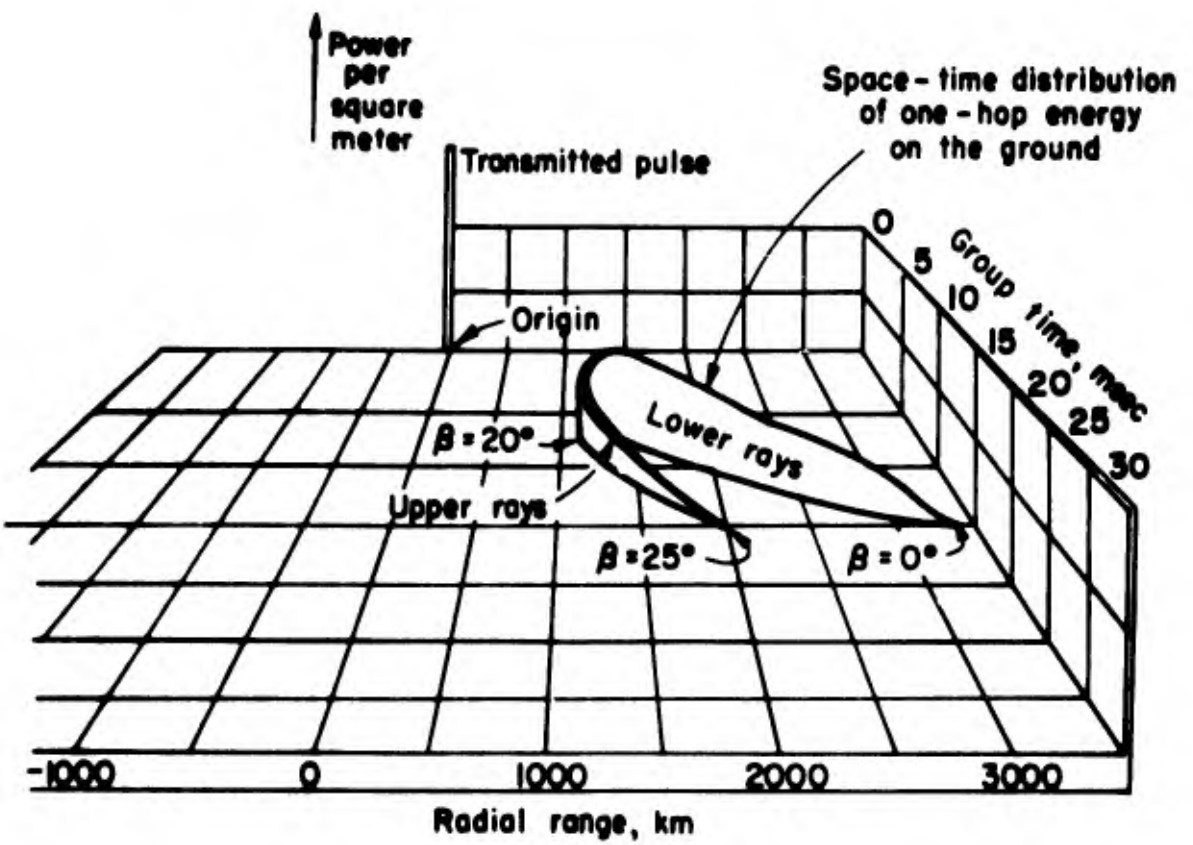


Fig. 16. THE DISTRIBUTION OF ENERGY IN SPACE AND TIME AFTER ONE HOP.

The returning energy is traced in Fig. 17 in a similar manner. The space-time distribution is similar but reversed because we consider only that energy which scatters back toward the radar. Finally, in

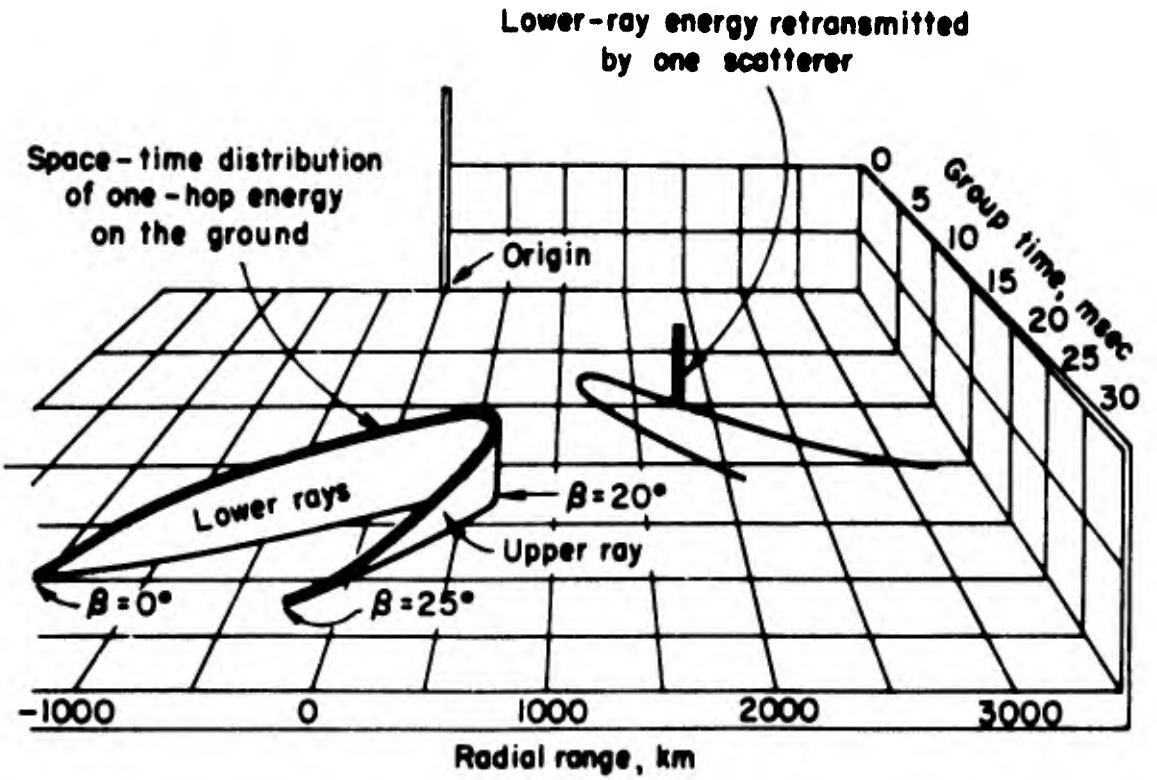


Fig. 17. THE DISTRIBUTION IN SPACE AND TIME OF THE ENERGY FROM A SINGLE SCATTERER.

Fig. 18 we see the contribution to the backscatter from a terrain strip of finite depth. All the dimensions of this figure can be evaluated in the computer, and so the terrain-strip concept of Fig. 18 is used as the integration step of a computer simulation process.

Initially, signal transmission time is computed to be very short. After the above integration process is completed, a numerical equivalent of the convolution process is used to account for the actual transmitted pulse duration and shape.

2. Backscatter Synthesis

The simulation process is difficult only because it involves a large amount of bookkeeping. Each mechanism which plays a part is fairly

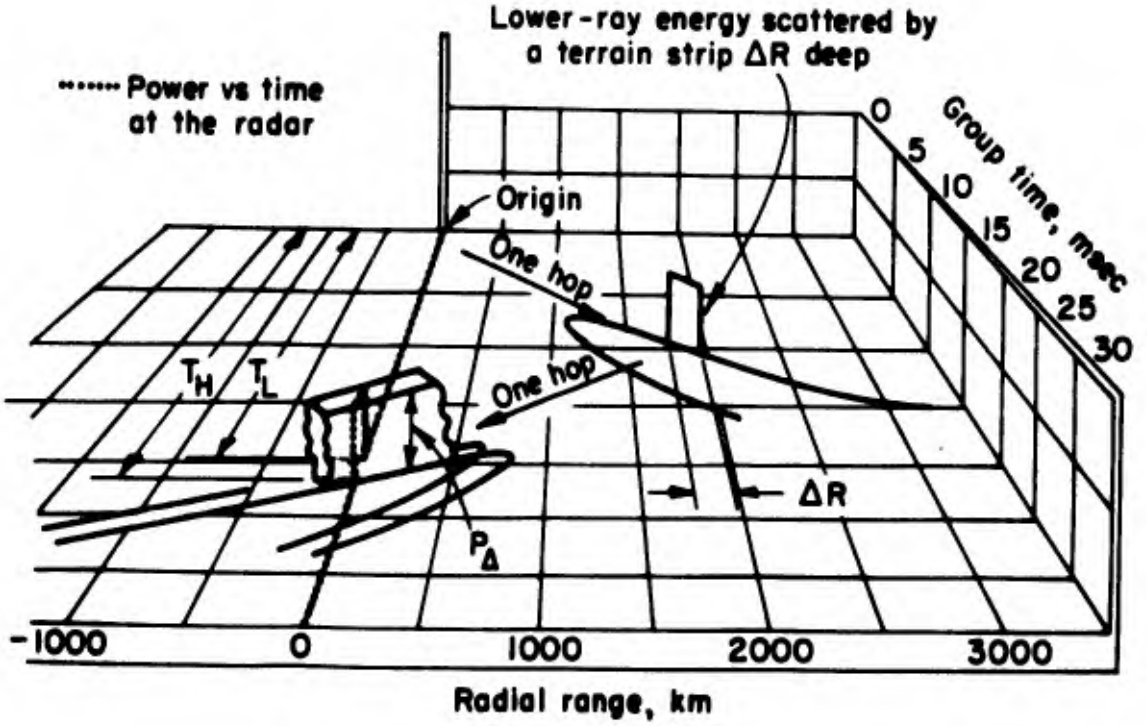


Fig. 18. THE TIME DISTRIBUTION OF ENERGY RECEIVED FROM ONE RANGE INCREMENT VIA SELECTED MODES.

well understood, with a possible exception of the actual scattering at the ground. The detailed spreading and folding of the originally hemispheric pulse can be calculated by raytracing. Absorption, ground reflection, ground scatter, antenna gain patterns, and the other factors which influence backscatter energy density can be accounted for in a straightforward manner, although the end result seems complicated because of the large number of interacting mechanisms. Finally, a computer can be instructed to find a family of fixed-frequency raypaths and the received backscatter for a given pulse shape, duration, time of day, etc. This can be done at many frequencies and thus a complete sweep-frequency backscatter record can be made synthetically by simulation of the natural process.

This backscatter amplitude calculation, like the raytracing itself, is a mixture of bookkeeping and evaluation of appropriate equations, and thus it is inherently well suited to a computer. When an integration of perhaps 6000 components is used to make a fixed-frequency

amplitude calculation, the process takes about 30 to 60 sec on an IBM 7090. An example of the resulting synthetic record is shown on Fig. 19. When calculations are carried out at different frequencies, a sweep frequency record can be generated as illustrated on Fig. 20. This shows a small three-dimensional model of the process; in practice, a special two-dimensional plot is used to represent this data display [19].

3. A Useful Approximation

If one is interested primarily in the gross shape of sweep-frequency backscatter, a much simpler technique can be used. Figure 21 shows an experimental record and a matching synthetic record. To make this synthetic record, the ionospheric structure shown as an inset in the synthetic record was used. Notice that electron density vs height curve is given at a number of different horizontal ranges measured from the transmitter. The difference between these curves shows the horizontal density gradients. After the rays were computed at several frequencies, a plot was made showing group delay vs frequency for every computed ray. At each resulting coordinate, a small horizontal dash was drawn. The density of raypaths on the ground is roughly analogous to the density of the energy on the ground, provided the rays are transmitted at equal increments of takeoff angle. It follows that the density of horizontal dashes on the frequency vs time graph is roughly a measure of the energy density of the backscatter.

The similarity between the experimental and synthetic backscatter records is readily apparent. This record was most striking because of the amount of structure present. For many ionospheric situations, the backscatter structure is much more simple than illustrated here; sometimes the structure appears hopelessly complex.

M. Simulation of Hypothetical Radio Systems

Because raytracing is essentially a simulation process, it can conveniently be used as a method for predicting the performance of hypothetical radio systems. It is usually cheaper to set up a computer simulation than it is to build a proposed system. The simulation also has an advantage in that the parameters of the hypothetical system can be easily

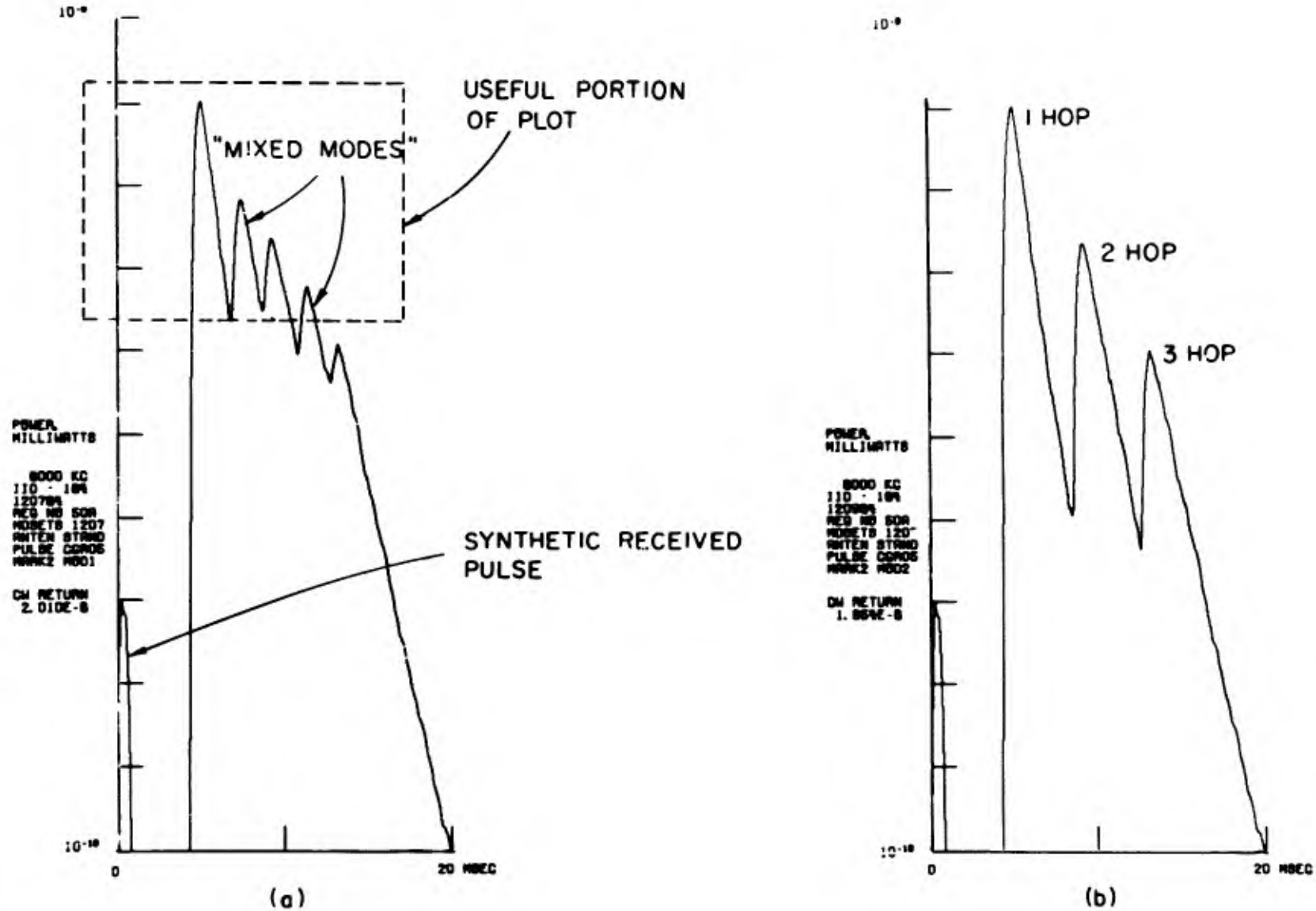


Fig. 19. SYNTHETIC BACKSCATTER SIGNAL STRENGTH PLOTS (a) WITH AND (b) WITHOUT MODE MIXING IN EACH INTEGRATION.

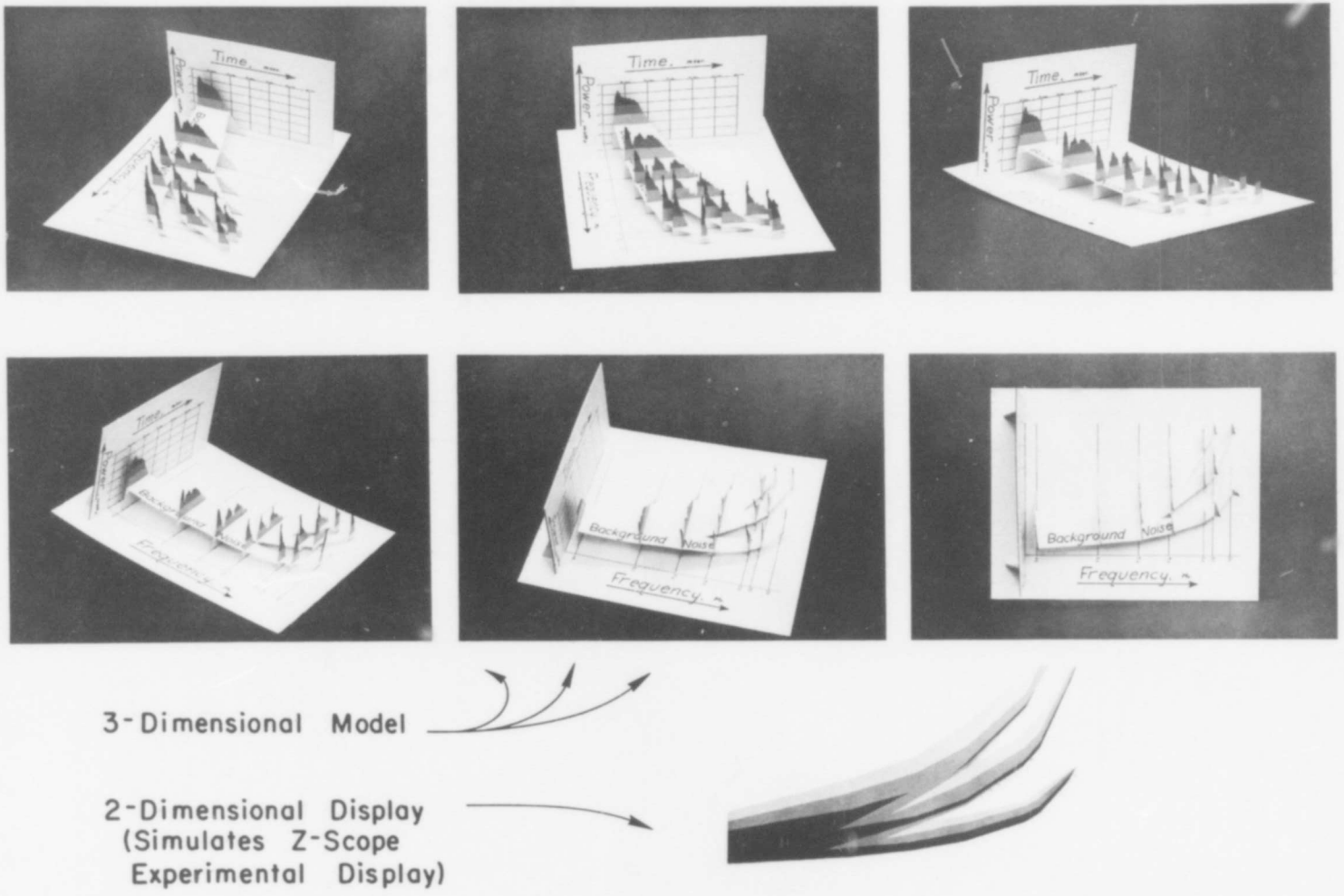
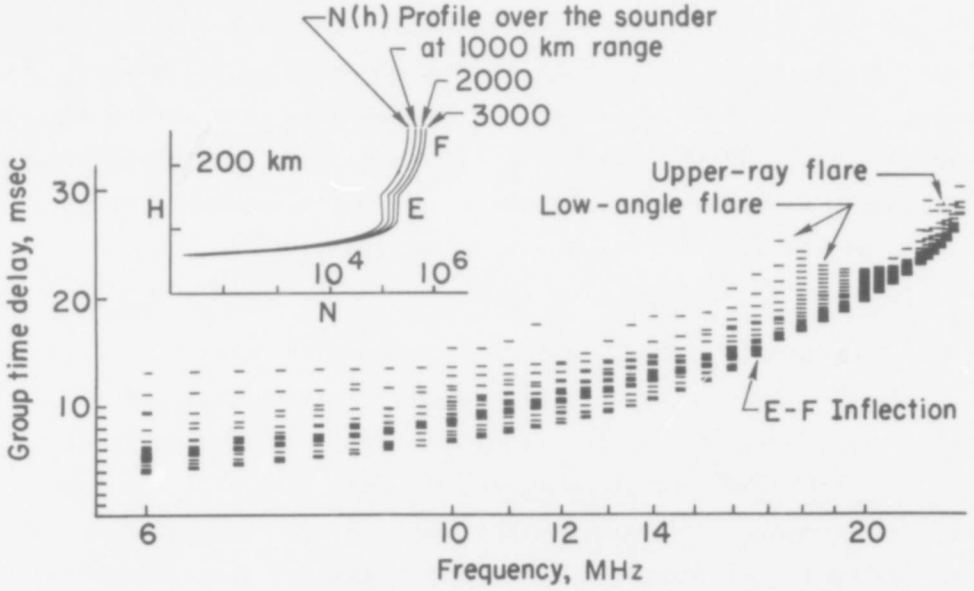
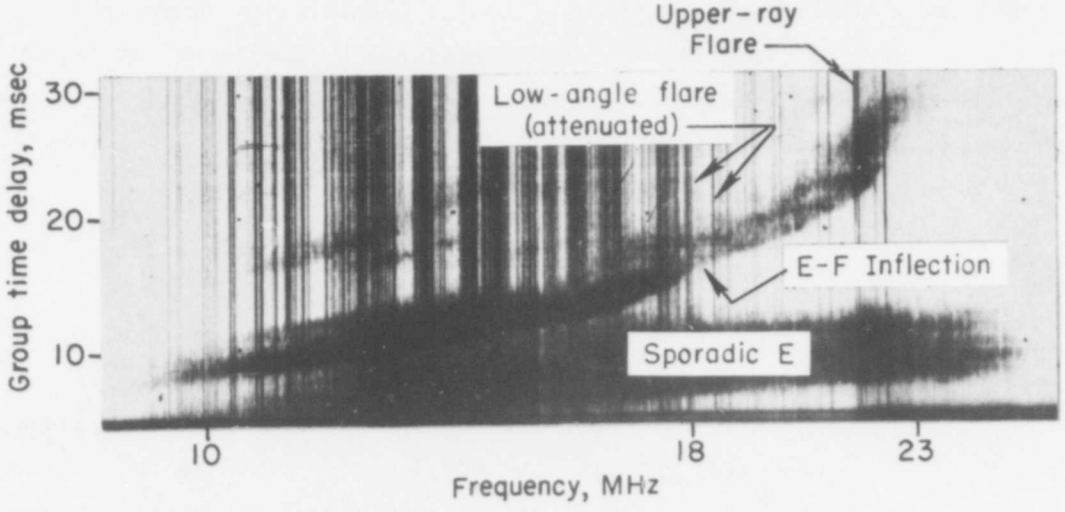


Fig. 20. A MODEL SHOWING THE ADDITION OF THE FREQUENCY DIMENSION.



(a) Computer plot



(b) Experimental record: 0855, 22 October 1964

Fig. 21. A SIMPLE METHOD FOR SIMULATING SWEEP-FREQUENCY BACKSCATTER STRUCTURE.

varied to find its optimum design. This application of raytracing has not been widely exploited, primarily because the simulation processes have only recently been developed and they have not yet earned the widespread confidence of the leaders in radio system design. Nevertheless, system design is probably one of the major future applications of raytracing. An illustrative example is given below.

During the simulation of an oblique ionosonde, it was noticed that the takeoff angle is a function of frequency which behaves in much the same manner as does the group delay. Since an ionogram is merely a display of delay vs frequency, it seemed reasonable to consider what could be learned about the ionosphere if one measured, instead, the takeoff angle vs frequency. To evaluate this proposal, a hypothetical sounding system called a "betasonde" was visualized (takeoff angle is called β) and the records which it might acquire were synthesized. Figures 11 and 12 show corresponding sets of real ionograms, synthetic ionograms, and synthetic "betagrams." Through the use of the rayset data, it was easy to make a synthetic betagram for each existing ionogram. Comparison of 135 pairs of such records showed many advantages and disadvantages of vertical angle sounding. For example, the technique might measure the direction of an ionospheric tilt, while the ionosonde is inherently insensitive to tilt direction because of reciprocity considerations which are unavoidable in measurement of time delay. The use of this simulation technique provided much information needed to evaluate the proposed sounding system.

N. One-Way Signal Strength Calculations

The most complex signal strength calculation yet encountered by the author was the ground backscatter routine previously described. In that case, the energy density was calculated outward to all locations on the ground and then backward from all locations on the ground to the radar. Often, however, one wishes to know the one-way signal strength along the ground from a single fixed transmitter in order to evaluate the effectiveness of the transmitter. Such calculations are easy to perform once raytracing of the path is completed, because all the necessary parameters are readily available.

For oblique propagation, the index of refraction stays roughly between 0.8 and 1.0 so that the deviative absorption is usually small, compared to the non-deviative absorption which takes place in the D region. If extremely accurate results are not required, one can compute a raypath and subsequently calculate the absorption by examining the orientation and location of the raypath when it encountered the D region. If better results are desired and if the collision profile is known, it is possible to compute absorption during the raytracing process, accumulating during each step the successive values in another "bookkeeping ledger." The choice between these two methods of computing absorption is again dictated by the quality of the data and the needs of the user. Note, however, that an improper collision profile can invalidate the more complex method [20], so it is not necessarily the "best way" to compute absorption.

Other signal strength factors are also readily available from the raypath computation. For example, the focusing due to the ionosphere is directly indicated by the spacing between adjacent raypaths. This "flux line" approximation breaks down at a caustic, and while the exact nature of the energy density variation in this situation is mathematically interesting, practical applications of raypath signal strength calculations are seldom hindered by this breakdown of geometric optics.

Also the ray approximation breaks down at the skip distance. The flux-tube concept leads to an infinite energy density at that point. Again, the problem is mathematically interesting and it has been solved by a full wave solution for a few special ionospheric structures [21,22]. The amount of energy involved in this (unreal) focusing is large if the electron density distribution is extremely smooth, as it usually is in the mathematical models such as Chapman and parabolic layers.

It is the author's impression that the importance of this problem has been overemphasized because of the use of unrealistically smooth ionospheric models. When electron distributions are measured by rocket probe, it is seen that the density is not nearly as smooth a function as are the mathematical models. Thus, in the real world, extremely high signals do not often arise because of skip distance focusing. In

pr? tical simulation by means of raytracing, the skip focusing can be handled by simple procedures which distribute the focused energy over a few kilometers (or if time is involved, over a few microseconds).

Within the framework described above, raypath calculation results can be used to calculate energy density manually, if only a few numbers are desired. When large amounts of data are to be processed, the rays stored for subsequent computer use in raysets can be used in a similar manner for signal strength calculations in the digital computer. The only essential difference between the manual and computer approach is the fast bookkeeping of the computer. For backscatter amplitude, the computer is needed to make the process practical because of the length of the calculation.

O. Measures of Raytracing Accuracy

One of the difficult questions which arises in raytracing is the determination and measurement of errors in the simulation system. The primary difficulty in this regard is the lack of knowledge of correct answers. The lack of a standard against which to compare makes it difficult to evaluate the performance of a particular raytracing technique. To the author's knowledge there are only two methods for checking accuracy.

1. Checks on Self-Consistency

Several things can be done to avoid major errors without having any other data to use as a check. The computing technique can be used to calculate a multitude of closely spaced raypaths in a very smoothly structured ionosphere; the resulting ray ranges should behave in a smooth manner. When this type of test is run with a computing technique which involves many integration steps, it is usually found that the range depends to some extent on the exact number of steps in any particular ray. If not corrected, such a calculation produces range vs takeoff plots which appear to be scalloped. Careful analysis of the cause of the individual scallops usually shows how to avoid this form of error.

Another self-consistency check is measurement of the symmetry of a ray when there is no magnetic field and no horizontal gradient of electron density. Under such circumstances, the ray apogee should occur at exactly half of the total raypath range and also the ray should land at exactly the same angle at which it left the ground. Failure to pass this type of test can usually be traced to some unsymmetrical condition in the individual integrations. When no magnetic field is present, ionospheric propagation (and its simulation) are governed by the principles of reciprocity and, as a consequence, the computational technique should produce the same answer even if the direction of the computation is switched from one end of the raypath to the other. The requirement for invariance under direction-switching applies to each of the integration steps also, and in some situations the symmetry is not easily achieved.

Other self-consistency tests are possible in addition to the two described above. For example, the length of the integration steps can be changed but the answer should not change. As integration steps are shortened, most computational methods achieve greater precision until the step gets so short that round-off error accumulates to equal or exceed the inherent systematic error of the technique.

2. Checks Against Error-Free Analytic Calculations

For some special ionospheric models consisting of a single layer, raypath parameters can be expressed analytically without any approximations [23,24]. The answers can be as accurate as one desires and are limited only by the number of significant figures available during evaluation of the expressions. When this technique is used, it is not unreasonable to expect answers which are correct to as much as 10 significant figures. Since typical raytracing techniques give answers that are correct to only 4 or 6 significant figures, it might be considered that these test ionospheres provide a "primary standard" for accuracy checking.

However, the exact analytic ray equation which are available suffer from two disadvantages. First, the raytracing technique under test must

be able to accept that particular ionospheric structure for which the analytic answers have been computed. This is sometimes impossible. Second, this method of testing provides no means for evaluating the ability of a raytracing technique to handle localized irregularities, geomagnetic effects, or absorption effects.

III. MAJOR PROGRAM TYPES IN USE

The author has striven to develop a variety of programs and to make each one available to other qualified organizations upon request. In order to keep track of program developments, we have adopted a system of identification similar to that used in the military services. Programs are divided into major families and the families are given "Mark" numbers. Within each family, new versions are assigned modification numbers or, more simply, "Mod" numbers. Thus each program is identified by a "Mark-Mod" number; for example "3, 19" means Mark 3, Mod 19. The major categories of programs are as follows.

Mark 1 programs are versatile and utilize Snell's law to calculate detailed ray trajectories as a sequence of short arcs. The ionospheres can be tilted, but this program makes an approximation which limits its usefulness to those ionospheres containing gentle tilts. (This limitation will be explained in more detail later.)

Mark 2 programs utilize analytic equations for ray trajectories in parabolic layers under the assumption that the entire ionosphere consists of no more than three concentric parabolic layers. This is an adaptation of the original program of Kift and Fooks [3] which has been further developed in this country [4].

Mark 3 programs are simplified for extremely fast operating using Snell's law in ionospheres which contain no tilts. This particular approach has been developed to an unusually high degree of efficiency, and it will be described in full detail in this report.

Mark 4 programs again use Snell's law, but they correctly account for the tilts in the gradient of the electron density in the ionosphere. This leads to considerable complication making the programs much more detailed than the similar Mark 1 programs. At the time of this writing, the computer programs existed but had not been fully debugged.

Mark 5 programs are designed to be used in conjunction with Mark 1. These programs use the output of Mark 1 programs as input, permitting a ray to be calculated in successive segments on successive computer runs. Thus a ray can be calculated through an ionosphere which has a more complex structure than can be conveniently described at one time to the computer. Other advantages exist when this form of operation is used.

Mark 6 programs compute sound rays in the atmosphere. One version includes the effect of winds aloft. By slight modification, these programs can be used to compute sound waves under the sea, including the effect of ocean currents. At the time of this writing, we do not have a working version of Mark 6 because our computer has been changed and we have not yet had the need to modify this program to make it work on the new facility.

Mark 7 (and all the rest of our programs) includes the magnetic field through the use of the Haselgrove equations. This particular program is a modified version of the one recently written by R. M. Jones [25]. As given to us, the program was extremely complicated and difficult to run because of the large amount of data needed in the set-up. Much of the complexity was eliminated in Mark 7 by restricting the ionosphere so that it could not have tilts. Like Mark 6, this program has not been modified for our new facility.

Mark 8 also includes the magnetic field and is a modified version of the program originally written by J. W. Finney of ESSA. We have changed it so that the input and output are similar to that of Mark 1 programs but otherwise it is much like the original. Electron density variation is only two-dimensional, but the rays leave the great circle plane because of the magnetic field action.

A. The Fastest Technique, Mark 2

When the only available information, concerning the nature of the electron density distribution, must be taken from the ionospheric

predictions, then some comparatively crude raypath assumptions can be justified because of the lack of detailed ionospheric information. In Mark 2, it is assumed that the ionosphere consists of a number of concentric parabolic layers, one above the other. In this model, it is possible to use analytic formulas for the path length of the ray which have been derived by Appleton and Beynon [26]. These formulas can be applied to the successive layers in turn, and thus a ray can be calculated from the ground up through the ionosphere and back again in only a few steps. As a consequence, these programs are extremely fast.

This technique was developed in several stages: First, it was discussed by Kift [2] and was subsequently implemented in a computer program by Fooks [3]. Later, the program was developed in the United States, at this laboratory, by Westover and Roben [4]. Finally, Westover modified it into the present Mark 2 program.

Because of the detailed description given to this technique by Westover and Roben [4], we will not repeat it here. A good deal of the programming is actually concerned with the generation of the ionospheric model from the data given by the ESSA predictions, used in conjunction with various empirical approximations which give the absorption, the maximum electron density in the E and F_1 layers, and the heights of the various layers. The semi-thicknesses of the parabolic layers are also fixed in accordance with empirical formulas. The entire procedure is a sequence of approximations, but nevertheless, there have been many useful applications of this program. Often, thousands of rays are needed and one cannot justify the expense of a very realistic simulation technique.

We have recently de-emphasized this program because a new approach taken by Neilson [27] appears to be better. For example in the Kift-Fooks calculation, a ray taking off at some angle, relative to the ground, will strike the ground after the hop with exactly the same angle, regardless of the ionospheric tilt. (It may miss the ground but, if it does strike it, the angle will be the same.) This approximation is unrealistic and is avoidable. Neilson's program accounts for the fact that a ray may gradually become shallower or steeper as it progresses from hop to hop over a long distance. We have not yet had an occasion to make any

quantitative test of Neilson's technique, compared to our current Mark 2 program, but the reader should be aware of these two similar approaches to the same problem.

B. Programs Utilizing Snell's Law, Mark 1, 3, 4, 5, and 6

In the ionosphere, the index of refraction depends upon both the direction of the magnetic field and the direction of the wave normal. Thus, the index cannot be described by a vector field, so tensor calculus is often used. In fact, it was through the use of tensor calculus that Haselgrove first derived his set of equations which are now so useful [6].

If the geomagnetic field can be ignored, the situation is greatly simplified because the index of refraction becomes a scalar. It can then be completely described by a scalar field, that is, a spatial distribution of a quantity which has a real value somewhere between 0 and 1 and which depends only on the position in space. There is no dependence upon the direction of anything. The simplification allows the use of Snell's law for the calculation of ray trajectories and is the fundamental reason why neglect of the magnetic field is worthwhile. Thus, a great increase in the computation speed and a consequent decrease in the cost are realized.

In the Mark 6 programs which compute sound rays, a similar simplification occurs when we neglect the wind. This is usually reasonable, although one version worked quite well when the vertical and horizontal components of wind (but not the rotational component) were included. These developments will be described later when Snell's law itself is discussed.

IV. ENCODING THE IONOSPHERE

The inventive process of writing a "raytracing program" actually consists of two discrete parts. The first part involves the encoding of the ionospheric model into functions and/or a table of numbers and the invention of a computer process for evaluating the functions and decoding the table of numbers to recover a close match to the original ionosphere. The second part is the raytracing itself, which must necessarily occur only in the decoded ionosphere.

It has been the author's observation that, in raytracing literature, the ionosphere coding aspect of raytracing has been unrealistically de-emphasized, possibly because writers do not consider it important. Nevertheless, the coding requires considerable effort and it is not always straightforward; there are good and bad methods in use. Unfortunately, the programmer usually does not have any basis for developing insight into these matters until he has obtained a working program and has tried to use it; only then does he find that his coding scheme is inadequate for the type of problems actually encountered.

In this chapter, various methods of ionospheric structure coding are described in an attempt to shed light on the strengths and weaknesses of each method. It is hoped that this will help others to intelligently choose among the existing methods or to invent new ones that are even better adapted to special needs.

All raytracing programs described here make use of a common basis for the description of the variation of electron density (or, in the case of waves, the description of the variation of sound velocity). As a mathematical matrix in which the coded data can be referenced, a system of concentric spherical shells is set up surrounding the earth. The fact that these shells are concentric with the earth sometimes leads to the mistaken impression that the ionosphere is not tilted.

Occasionally, the ionospheric structural information, available from outside sources, consists of a series of heights that are associated with a series of electron density (or plasma frequency) values. For example, one might be given data telling the height vs distance of the line along which the electron density is 10^5/cc. When this data is given, it is tempting to program the computer so that the electron density is specified by a height vs range of the fixed density. This method offers a slight advantage in the calculation of the angle of the density gradient which is simply perpendicular to the isodensity line. However, there is a serious disadvantage to this method of programming because it is awkward to insert a new isodensity line beginning at some range other than zero. It is also difficult to stop an existing isodensity line at a range short of the maximum. In the ionospheric models which are interesting, there is frequently a closed isodensity line which exists only within a small region. While it is not impossible, it is nevertheless very difficult to program such an electron density model through the use of descriptions of the height-range locus of an isodensity line. Because of this difficulty, we have chosen to avoid the programming technique.

All of the programs at this laboratory, whether invented here or invented elsewhere and modified here, are designed to use electron descriptions given at a series of fixed heights. In describing these coding methods, there is a possibility of semantic confusion between "range" measured along the ground and "range" measured as a straight line between some observer and a distant observed point. In the following discussions, the word "range" and its symbol, R, are meant to signify the great circle distance measured along the surface of the earth from some origin (which will be understood) to some point on the surface of the earth, located directly beneath the point being described. This difficulty could probably be avoided if we had a convenient word for the geocentric angle subtended by the origin and the described point. Here, this geocentric angle will be called θ. The following discussions will use range R in this special sense, so that $R = 6370\,\theta$ always, where 6370 is the standard earth radius in kilometers [28].

A. The NR^2 Code

This code is designed for use when the available ionospheric information consists of two or three vertical ionograms taken from different sounders located along the path of interest. In this circumstance, we know the location of the simulated HF transmitter or receiver that will be the site from which rays are to be computed. We also know the vertical profile of electron density vs height above the two or three sounder locations which hopefully lie along the great circle in which propagation is to be studied. It may be that one of the sounders is behind the HF station and the others are ahead of it, or all sounders may be ahead of the HF station. First, we select a series of heights at which we wish to describe the electron density above each of the sounders. Next, we find the appropriate lists of electron densities obtained from the reduced ionograms at the selected heights.

One must then select a method for calculating the electron density at some specified height between the sounder profiles. There is no information available to tell the right answer. If there are only two soundings available, then the author usually chooses linear interpolation in electron density vs range at each altitude. If there are three soundings available, then we sometimes choose to use the NR^2 procedure which fits a second-order curve of electron density vs range to the three data points at each discrete altitude. Mathematically, the NR^2 procedure is as follows:

$$\text{At each height,} \qquad N = N_o\left[1 + k_1(R-R_o) + k_2(R-R_o)^2\right]$$

Here, N_o, k_1, and k_2 may be different at each height. They are selected so that N_o is the electron density at the selected height above the first sounder on the path (the one behind the HF facility, if this is the case). At the first sounder, $R = R_o$. Similarly, k_1 and k_2 are computed by matrix inversion so that the expression for N has its correct value when the range R is that of each of the other two sounders. Beyond the last sounder (let us say at range R_2), the electron density remains unchanged as a function of distance, having a value

$$N_2 = N_o\left[1 + k_1(R_2-R_o) + k_2(R_2-R_o)^2\right]$$

The advantage of this second-order coding method is that the electron density is a smooth function of distance within the span of the sounders. Such smoothness is more likely to be representative of the physical situation, and so in a sense, we make the best use of the meager data by adopting this code. However, the method contains a trap for the unwary and must be used with some caution for the reason outlined in Fig. 22. If there are two close sounders and one more distant, then it is possible that the electron density coded in the NR^2 method will have a very unlikely value after decoding. This is because we force smoothness to second order using only three points, and the fitted curve can make an excursion far from the average value of the input data in the region between the distant sounder and its neighbor. Such a form of error can also arise with uniformly spaced sounders but it is less likely. To avoid such a circumstance, one should always plot the three input N vs h profiles above the sounders and examine them to see that the intermediate points will be logical. If there is any question, we compute and plot N vs H from the formula at a number of intermediate ranges to look for unintended effects. Thus it is seen that the NR^2 method should be used only by someone familiar with ionospheric structure.

B. Linear Interpolation

This coding method is probably the most foolproof, but it introduces unrealistic discontinuity in the density vs range function. The N vs h listing is obtained for a number of different sounders or from the predictions. The number of N-h profiles is limited only by the patience of the user and the memory size of the computer. There is no curve-fitting; so there is no limit on the number of profiles. Between each profile at each height, the electron density is assumed to vary linearly with respect to range. It is difficult to imagine a simpler coding scheme and thus the method is well suited for situations where the raytracing personnel are comparatively inexperienced. (At the cost of a slight additional loss of

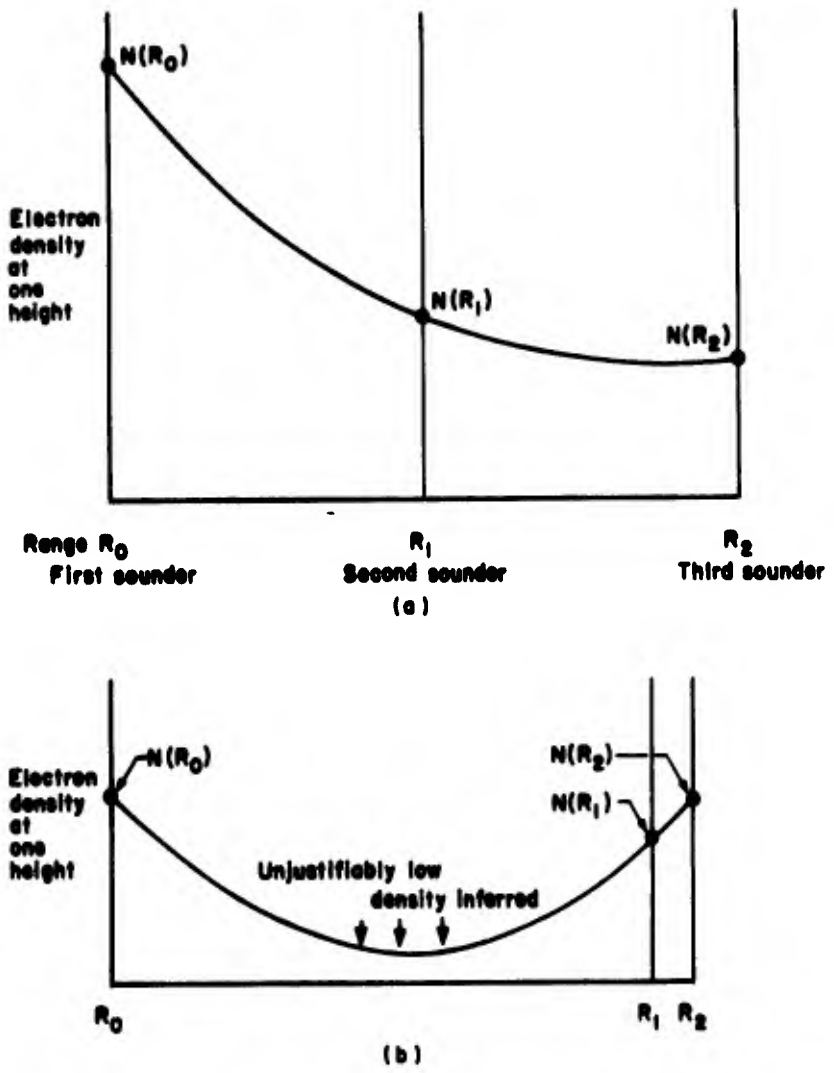


Fig. 22. ELECTRON DENSITY VS RANGE WITH THE NR^2 CODE. (a) Electron density vs range at one height when sounders are at near-equal spacings and (b) an illustration of the difficulty that could arise if the NR^2 code is used when two of three sounders are close ($R_1 \approx R_2$).

realism, one might interpolate in refractive index vs range. This might be more economical in some applications.)

In order to make the computer programming reasonably tractable, we impose a requirement that each N vs h profile must contain one point for each chosen height, so that all N-listings are of equal length. The

specific method found to be useful for computing oblique HF rays is the following: Beginning at a 60 km altitude in the daytime or an 80 km altitude at night, we make one N point for each 1 km of height up to 300 or 350 km, the upper bound depending on the data. Then all the points are placed on one punched card per height. As it turns out, it is reasonable to put ten N points and one h value on each card; this permits ten soundings to be programmed into a single deck of about 300 cards. Of course, it is a rare path for which ten soundings are available, but when the ionosphere is taken from ESSA predictions one may easily choose ten spots along the path at which vertical profiles are to be coded.

C. The "Shape" Code

Frequently one wishes to compute rays through an ionosphere in which there is a localized irregularity of some particular shape. For example, this author has worked with traveling gravity waves in an otherwise undisturbed ionosphere and also with local models of chemical releases. For the programming of such anomalies, hereafter called "shapes," we have found it adequate to use the following minor variation of the NR^2 method: The locations of the beginning and end of the second-order expression are made to be functions of height. Thus where we once had three constants characterizing each chosen height, we now have five constants. Since the height is also coded on the punched cards, each card now must contain six numbers. Mathematically, the system is as follows:

At height m, the values of N_{mo}, k_{m1}, k_{m2}, R_{mo}, and R_{m2} are uniquely assigned, independent of the corresponding values at any other height.

For $R < R_{mo}$, $N = N_{mo}$

For $R_{mo} < R < R_{m2}$,

$$N_m = N_{mo}\left[1 + k_{m1}(R-R_{mo}) + k_{m2}(R-R_{mo})^2\right]$$

For $R > R_{m2}$, N is constant, having the value

$$N_{m2} = N_{mo}\left[1 + k_{m1}(R_{m2}-R_{mo}) + k_{m2}(R_{m2}-R_{mo})^2\right]$$

It can be seen that the process outlined above establishes the outer boundary of the shape, matching the boundary description with the description of the ambient ionosphere. After boundary matching, there still remains one degree of freedom at each height, since we have specified only the electron density at the endpoints of each parabolic arc but we have not yet chosen the central value of the arc. (This corresponds to the central sounder of three in the NR^2 code described previously.) To finish the process, we usually choose the electron density vs height along the centerline, half way between R_0 and R_2, setting it to equal some constant value, or to some percentage of the ambient, or according to some analytic formula which fits the specific application.

To illustrate the application of the shape code, a small Plexiglas model was assembled. It was a three-dimensional device showing electron concentration vs altitude vs range. An overall view of the assembled model is given in Fig. 23. This model represents a circular disturbance (in altitude vs range) within which the electron concentration decreases to about 10 percent of its ambient value at the center. The model could be partially disassembled in a method analogous to the separation of a deck of computer input cards which represent the ionosphere. This analogy is appropriate because each input card represents electron concentration vs range at one single altitude; thus, the information on a single card is equivalent to that contained in a single model sheet which lies at a fixed altitude. On the model, there are eleven sheets, one for each 10 km from 150 to 250 km. These features of the model are most easily seen in its partially assembled state as shown on Fig. 24. Part (a) of the figure shows the parabolic electron density vs range at two different heights. (There is a third parabolic arc on the Plexiglas sheet which lies vertically over a range of 50 km; this was needed for mechanical stability and does not play any part in the analogy under discussion.) Figure 24 shows five stages of assembly, wherein the only change between the pictures is the addition of more levels to the model.

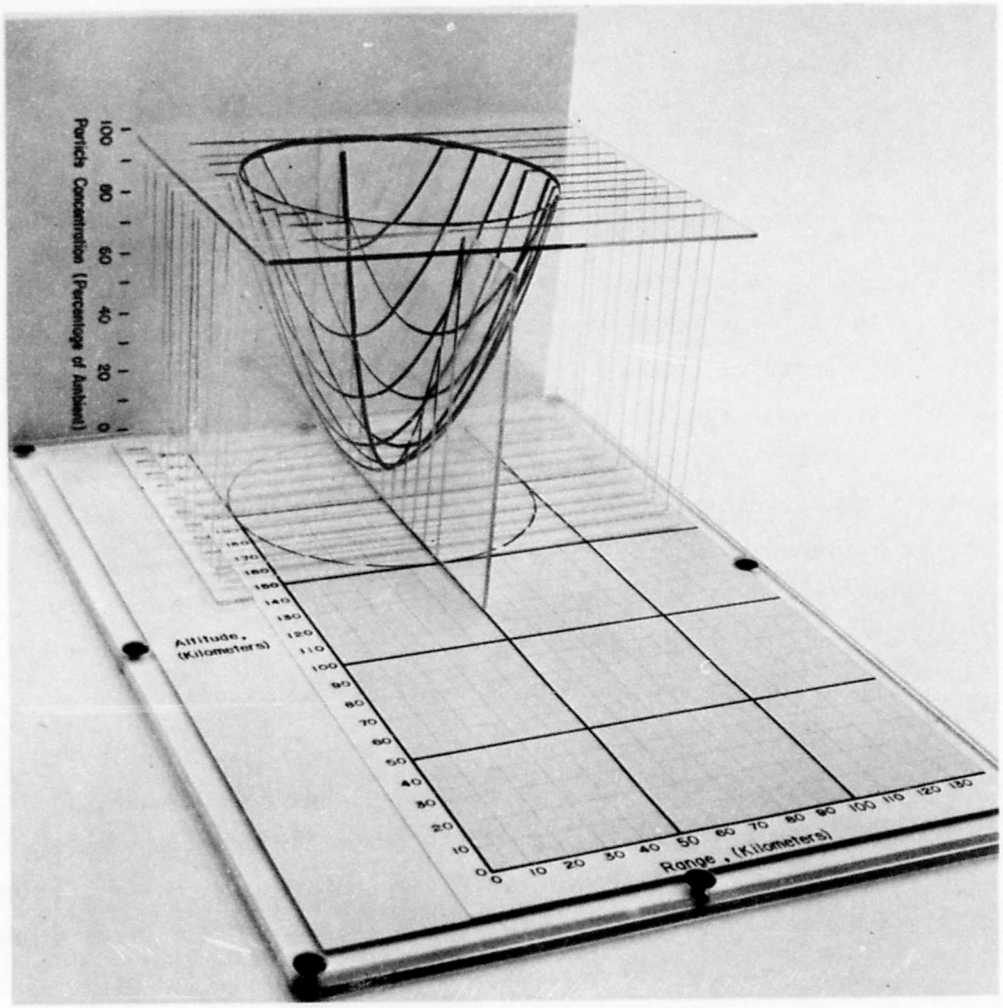
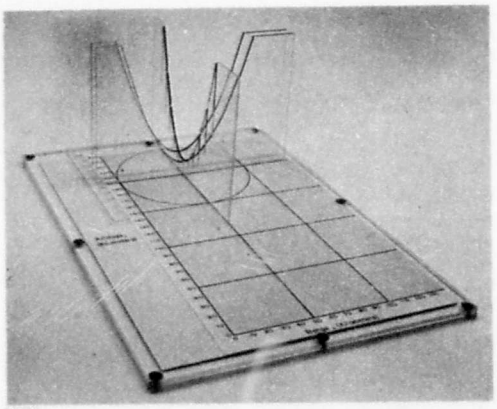
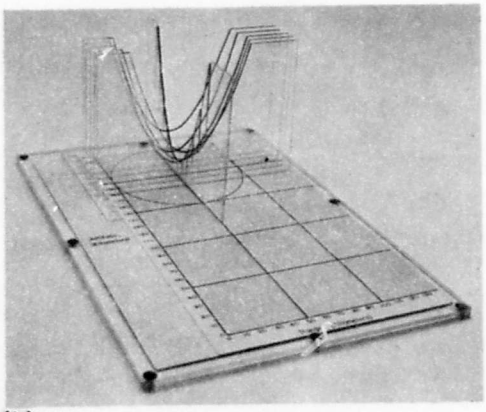


Fig. 23. MODEL ILLUSTRATING THE "SHAPE CODE" FOR CODING LOCALIZED IONOSPHERIC DISTURBANCES.

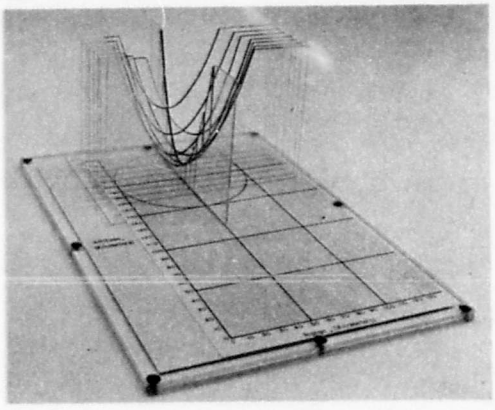
Figure 23 then shows the assembled model with an extra flat sheet laid on top to emphasize the circular outline of the whole structure. In the digital computer code, we would have a parabola every 1 km instead of every 10 km as illustrated in the model, but it was not practical to construct this entirely. However, one can see that, if we did indeed have a Plexiglas parabola for each kilometer of height, the circular outline would be obvious without the necessity for emphasis provided by the top plate in Fig. 23.



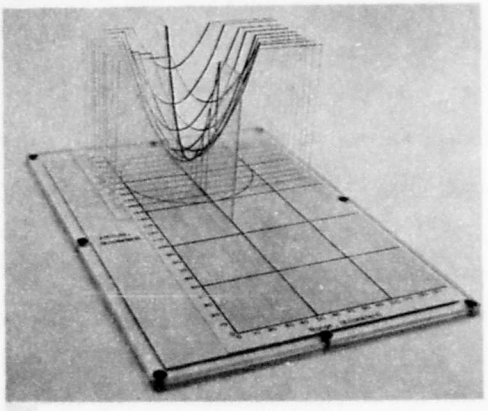
(a)



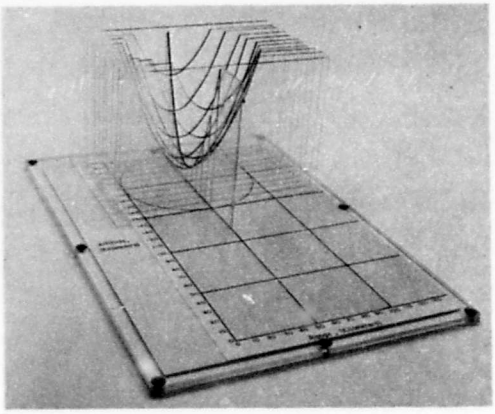
(b)



(c)



(d)



(e)

(a) Levels 200 and 210;

(b) levels 180, 190, 200, 210, and 220;

(c) levels 170, 180, 190, 200, 210, 220, and 230;

(d) all the parabolic levels;

(e) all levels below 150 and above 250 have constant electron density and the models are rectangular sheets such as the two illustrated.

Fig. 24. STAGES IN THE ASSEMBLY OF THE MODEL.

Each Plexiglas sheet then contains a parabola, and each such parabola is described in the computer program by four constants: k_1, k_2, R_0, and R_2. A fifth constant on each card gives the height and a sixth gives the ambient electron density. This is not illustrated on the model which simply shows the particle concentration as a percentage of ambient. This model may be considered as describing a function of altitude and range which is used to multiply an otherwise nontilted ionospheric model.

D. RTW Code

When it is desired to compute rays around the world (RTW), the process is simplified if the electron density model is periodic in a distance equal to the circumference of the earth, roughly 40,000 km. This can be accomplished by an easy modification of the NR^2 code wherein the second-order function is replaced by something involving sinusoids of a 40,000 km period. For example, one might choose to use:

$$N = N_o\left[1 + k_1 \cos (k_2 R)\right]^m$$

where k_2 is chosen to set the 40,000 km period and m is chosen to shape the function in some desired way. Then N_o and k_1 would be chosen so that $N_o(1 + k_1)^m$ is the proper daytime N vs h profile and $N_o(1 - k_1)^m$ is the nighttime profile. Other similar modifications can be readily composed. The changing of the computer program to accommodate this variation is little more than the changing of one card, the one originally containing the function $N = N_o(1 + k_1 R + k_2 R^2)$. It is not even necessary to change the data format on the ionosphere coded deck of cards since the numbers are of a similar nature.

E. Additive Models

All of the preceding codes share the trait of their electron density being found by multiplying functions. It is also practical to add functions with very little change in programming methods. Recently, Neilson [27] found it necessary to use addition, to permit the application of a special analytic approach to the ray calculation. Nevertheless, to this

author's knowledge, most applications of raytracing in tilted ionospheres have made use of multiplicative processes, probably because of considerations related to the physical processes that cause the tilts. For example, the passage of an atmospheric gravity wave influences the electron density by some percentage, not by some fixed number of electrons.

One cannot say in advance whether additive coding will be better or worse than coding that uses multiplication. Fortunately, the switch from one to the other is an easy programming task, so it is not necessary to commit oneself permanently to one of these two options.

F. Combinations of the Codes

The method of code interpretation inside the computer is such that it would be practical to use combinations of the methods outlined above, although the author has not yet found it necessary to do so. For example, suppose one had ionospheric information which permitted deduction of the widespread, long-enduring tilts in the ionosphere. This "ambient ionosphere" could be encoded by the linear interpolation logic. If one wished to simulate the passage of a gravity wave through this tilted ionosphere, then the shape code could be used to describe the localized disturbance as a percentage change from the tilted ambient. The simultaneous application of two multipliers (according to the logical controls of each) is a straightforward application of a digital computer. This combination has not been implemented primarily because of the lack of ionospheric information, not because of difficulty in the processing.

G. The Source of the Common Number, 80.6

In converting from electron density to refractive index, one uses the common relation $n^2 = 80.6\ N/f^2$, where n is refractive index, N is electron density, f is radio frequency, and MKS units are used. On occasion, the author has seen documents in which others have used 80.5 instead of 80.6. The parameter results from an evaluation of the expression $e^2/4\pi^2\ \epsilon_0 m$ whose derivation can be found in most texts on ionospheric radio propagation. The value of ϵ_0 is known, but the values of e and m are uncertain. The author examined various physical tables

which purported to be modern and found, for both e and m, four different values roughly centered about $e = 1.602 \times 10^{-19}$ and $m = 9.108 \times 10^{-31}$. The four listed values of each parameter can be used to derive 16 parameter values. This calculation is carried out in Table 1, and it can indeed be seen that 80.6 is considerably better than 80.5.

Table 1

EVALUATION OF THE CONSTANT $e^2/4\pi^2 \epsilon_0 m \approx 80.6$

e =	m =			
	9.10917	9.10891	9.10746	9.10696
1.601889	80.59	80.59	80.61	80.61
1.601839	80.59	80.59	80.60	80.61
1.602073	80.61	80.61	80.62	80.63
1.602117	80.61	80.62	80.63	80.63

Note added in proof: The April 28, 1969 issue of a magazine called "Scientific Research" lists new values of e and m attributed to Parker, Taylor and Langenberg with the disclaimer that they are "subject to change until formal publication". The new values are $e = 1.6021917 \times 10^{-19}$ and $m = 9.109558 \times 10^{-31}$. Using $\epsilon_0 = 8.854 \times 10^{-12}$ farads/meter, the constant becomes 80.62.

V. THE USE OF SNELL'S LAW TO COMPUTE RAYS

When we use the term "Snell's law," many people interpret this to mean only the well-known formula, $\mu \sin \Phi = \mu' \sin \Phi'$. While it is true that this is one form of Snell's law, we shall see that there are several other useful forms which bear little superficial resemblance to this; nevertheless, they are fundamentally the same. It might, therefore, be worthwhile defining "Snell's law" as it will be used in this report.

The index of refraction is a measure of wavelength at a point. Changes in wave direction are brought about by spatial variations in wavelength, just as changes in the direction of a line of marching soldiers are caused by varying the step length along the line of men. The analogy is most apparent if we choose to examine the cause of refraction by using Huygens' wavelet construction to find the successive locations of a wavefront at successive instants.

Of course, there is a difficulty in this analogy because individual soldiers in the marching line may not walk in the direction perpendicular to the formation. In wave propagation, an analogous difficulty arises when the refractive index is anisotropic; in this case, the energy of the wave may not propagate in the direction perpendicular to the wavefront.

In raytracing, we wish to track the progress of energy. For our purposes here, a ray is considered to be the locus of a tagged bundle of energy. (Much has been written about the validity of this concept, but it will be accepted here without further comment.) If there is a magnetic field, then the index of refraction is not isotropic and the ray and the wave normal are usually not aligned. Furthermore, the refractive index depends on the direction of the wave normal. If we trace a ray from one point to the next, then at the second point we know only the ray direction. In order to compute the wave normal direction, we have to compute the index of refraction but, in order to compute the index of refraction, we have to know the wave normal! Seemingly, this is an impasse. In fact, this difficulty is the specific reason why it has not been found practical to use the simple forms of Snell's law for raytracing in an ionosphere including the magnetic field. (Other methods do work; for

example, the set of linear simultaneous differential equations derived by Haselgrove can be used to find raypaths in anisotropic media.)

Avoiding the difficulties outlined above, we will use the term "Snell's law" in this report to signify the governing law for ray refraction in an isotropic medium where the wave normal and the ray are collinear. For ionospheric radio propagation, the consequence of this definition is that we must neglect the magnetic field. It will be shown that this neglect is reasonable for frequencies above about 6 MHz and that the degree of error introduced by the approximation decreases with increasing radio frequency. Thus, for the remainder of this chapter, it will be considered that the refractive index varies with position in space but not with the direction of anything; it is a scalar field.

A. Forms of Snell's Law

Let us examine the various ways in which Snell's law can be stated mathematically. To begin, we will derive the most familiar form.

1. The Trigonometric Form

It is the author's impression that this form of Snell's law is widely recognized because it is introduced in the education process at such an early point that the teachers must assume the students are ignorant of the principles of calculus. In a way, this is a pity for it will be seen that the other forms have utility and can be used to gain insight not easily gleaned from this trigonometric equation.

Figure 25 shows the conventional drawing, with a wavefront on one side of an interface and a later wavefront in the same wave train on the other side of the interface. The refractive index is assumed to be constant on each side of the interface with a sudden change in value at the interface.

In order to keep this derivation general so that it may also be applied to sound waves, we will use phase velocity instead of refractive index, recognizing that for radio waves the phase velocity v equals the speed of light c divided by the refractive index μ. For convenience of notation, it is assumed that unit time is required for the wave

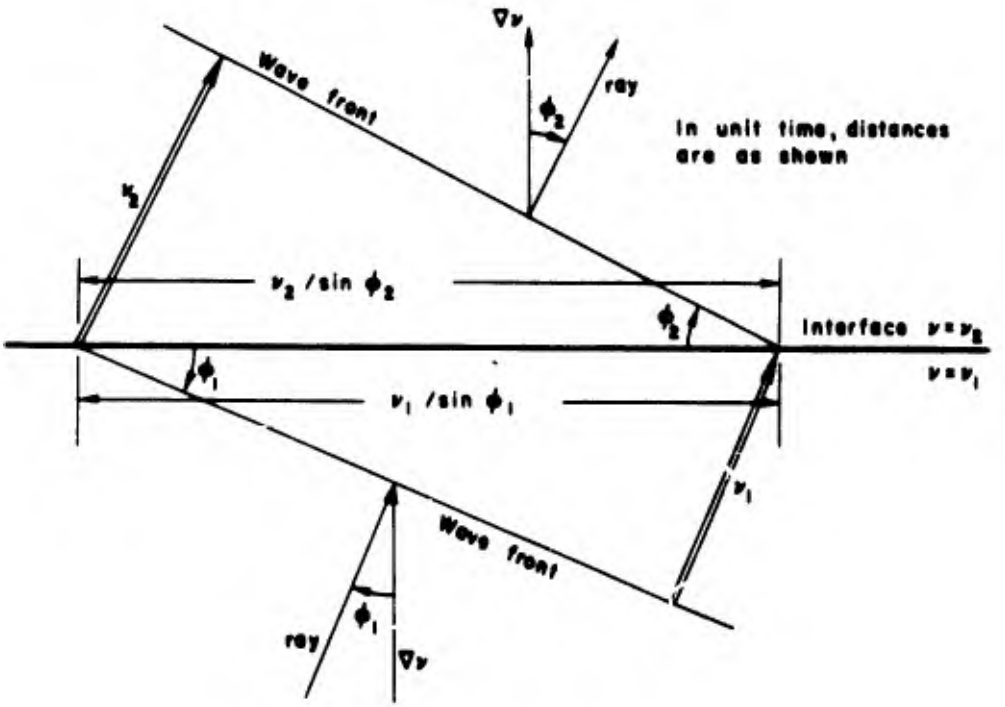


Fig. 25. GEOMETRY UNDERLYING THE SINUSOIDAL FORM OF SNELL'S LAW.

to travel between the two positions shown in Fig. 25 so that time does not appear in the equations. This also simplifies the later derivations.

The gradient of the propagation velocity, ∇v, is necessarily perpendicular to the interface. It is conventional to define the incidence angle of the ray, φ_1, as the angle from the gradient to the ray. Similarly, the transmission angle is defined as φ_2, shown on the figure. It can be seen that the distance along the interface may be expressed in two ways:

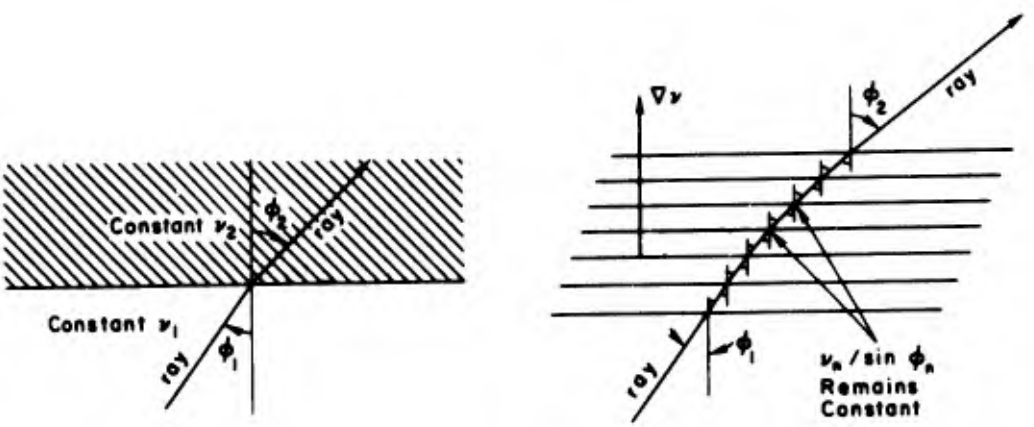
$$v_1/\sin \varphi_1 = v_2/\sin \varphi_2$$

For radio rays, phase velocity $v = \frac{c}{\mu}$. Making this substitution,

$$\frac{c}{\mu_1 \sin \varphi_1} = \frac{c}{\mu_2 \sin \varphi_2} \quad \text{or} \quad \mu_1 \sin \varphi_1 = \mu_2 \sin \varphi_2$$

These are the familiar forms of Snell's law.

In the derivation above, the sinusoidal form was calculated across a single discrete boundary between two homogeneous media, as sketched in Fig. 26a. However, the equation applies equally well to any continuous medium in which the gradient of phase velocity has a constant direction. As illustrated by Fig. 26b, this medium can be treated as a succession of discrete boundaries between an infinite number of layers having homogeneous media but infinitesimal thickness. At the first boundary encounter, $v_1/\sin \phi_1 = v_2/\sin \phi_2$. At the second boundary encounter, $v_2/\sin \phi_2 = v_3/\sin \phi_3$ but this, in turn, must also equal $v_1/\sin \phi_1$. This step can be executed any number of times and the equality will still hold. The concept illustrated in Fig. 26b is very important because it means that $v/\sin \phi$ is conserved (i.e., its value never changes) along a ray which traverses any medium in which the gradient of phase velocity has a constant direction. Mathematically, this occurs in Cartesian coordinates when the phase velocity (or the electron density or the refractive index) is a function of only one dimension. This is often assumed to be the case in analysis.



a. A single application of Snell's law

b. A repeated application

Fig. 26. REPEATED APPLICATION OF SNELL'S LAW SHOWING THE INVARIANT PARAMETER.

2. Bouger's Rule

When a ray traverses a medium in which the refractive index has circular or spherical symmetry, then a special form of Snell's law can

be derived that is closely related to the common form given above. Figure 27 shows the geometry of a ray which passes through the top and bottom surfaces of a homogenous slab circularly disposed about the center of the curvature shown. Because the slab is homogeneous, the ray is straight while it is within the slab. The refractive index jumps to a new value on either side of the slab so that the previously derived form of Snell's law can be applied at points b and c. The two points and the center of curvature define a triangle in which the law of sines requires

$$\sin \varphi_c = \sin \beta_b \frac{\rho_b}{\rho_c}$$

Furthermore, at point c the trigonometric form of Snell's law requires that:

$$\sin \varphi_c = \sin \beta_c \frac{\mu_c}{\mu_b}$$

where μ_b and μ_c are the refractive indices at points b and c, respectively. Equating the right sides of these two equations and rearranging,

$$\mu_b \rho_b \sin \beta_b = \mu_c \rho_c \sin \beta_c$$

which is Bouger's rule.

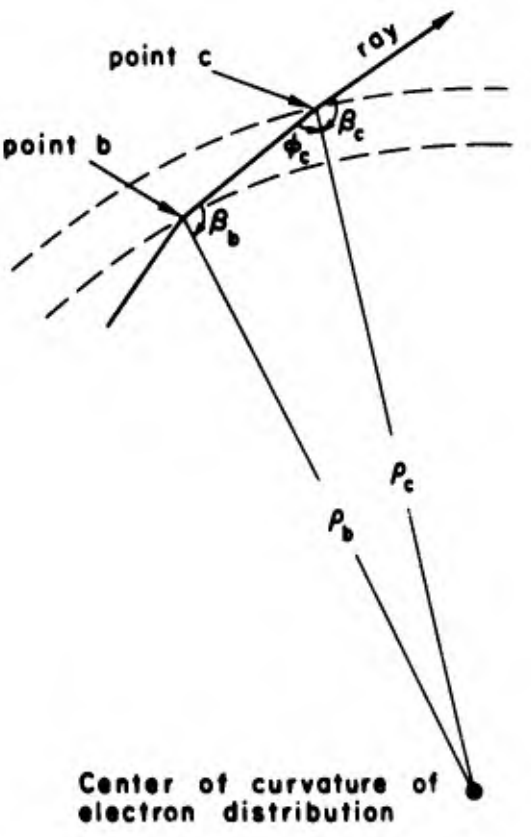


Fig. 27. THE GEOMETRY USED TO DERIVE BOUGER'S RULE.

Notice the similarity of this equation to the common form. It can be seen that the ray could be followed through an infinite number of successive points like b and c and the product $\mu \rho \sin \beta$ would be conserved, just as the product $\mu \sin \varphi$ was conserved in the Cartesian case. This is a more useful way to think of Bouger's rule. The conservation of the product provides a convenient basis for raytracing in a nontilted, spherical ionosphere. The rule can be applied to the ray as

it leaves the earth at a known takeoff angle where the product equals simply the sine of the takeoff angle multiplied by the radius of the earth. The author chooses to call this product "Bouger's constant," although the reader should be advised that this is not common practice. Bouger's rule serves as the basis for very fast Mark 3 raytracing program which will be described later in this chapter.

The author is not aware of the original reference for Bouger's rule. Dr. John Kelso of ITT-EPL has suggested that it may be derived from the work of the French scientist Bouguer. If this is so, then the name is misspelled throughout this report.

3. Snell's Law in a Tilted Ionosphere

It has recently come to this author's attention that Bouger's rule can be extended to an ionosphere which is not spherically symmetric. In 1963, the Russian, Ya. L. Al'pert, wrote a paper which has been translated into English [29]. He derived an equation which is, in the notation of this report,

$$\mu\rho \sin\beta = \text{Bouger's Constant} + \int_{6370}^{\rho} \frac{\partial\mu}{\partial\theta}\, dS$$

where S is the path length. We are currently investigating the implications of this equation to see if it can be used as a basis for efficient raytracing in a tilted ionosphere. At first glance, it looks quite promising.

4. A Useful Differential Form

Let us assume that the refractive index varies only in two dimensions so that a beam of rays in the plane of variation will never leave the plane. This differential form can be readily derived without the two-dimensional approximation (see, for example [1]). The two-dimensional case will be used here, however, because it appears to the author that this form provides the clearest insight.

Figure 28 shows a beam of rays traversing the medium with a wavefront drawn at two successive locations considered to be separated by infinitesimal times and distances. In the region containing the two

wavefronts, it can be assumed (because of the infinitesimal qualification) that the beam forms a circular arc about some center of a curvature as shown. Let the phase velocity at the inner edge of the beam be v. Then, if we move toward the infinitesimal distance $d\rho$, the phase velocity must be $v + dv$. From the geometry shown, the following equations can be derived:

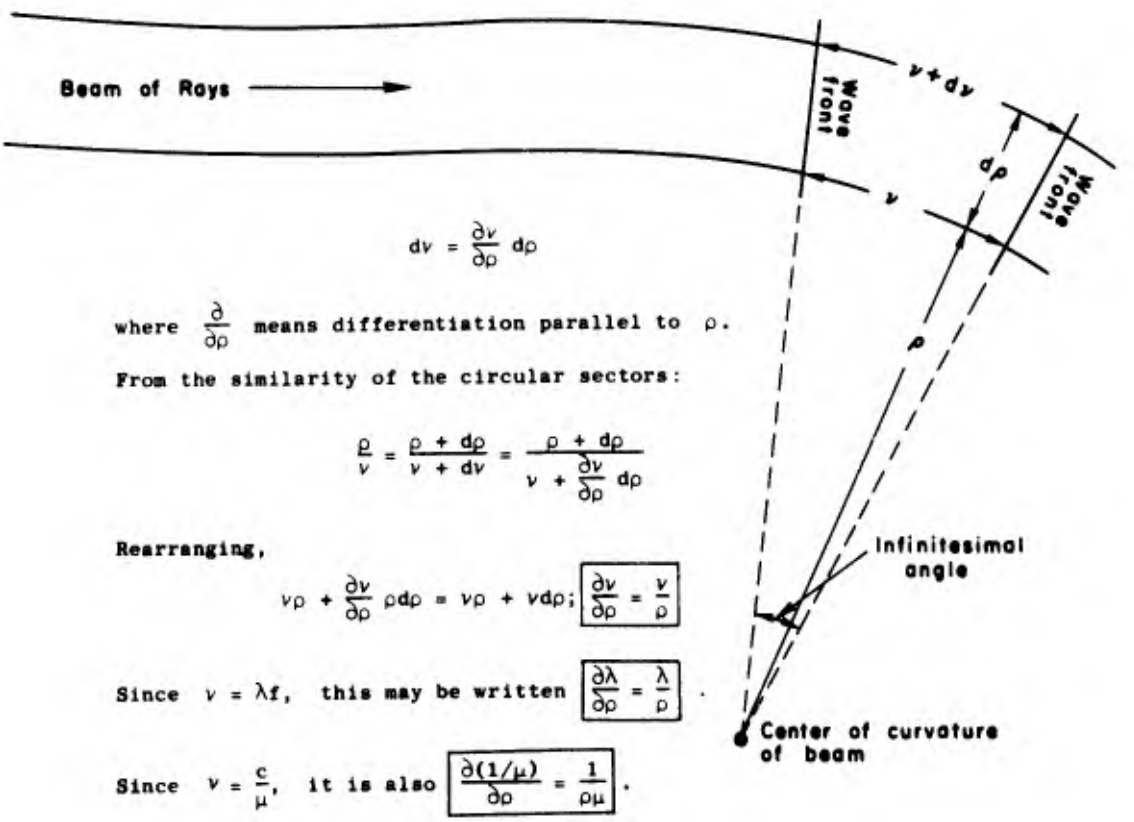


Fig. 28. DEVELOPMENT OF THE DIFFERENTIAL FORM OF SNELL'S LAW.

The last three equations derived are Snell's law. There is a symmetry of form in the first two equations which is lacking in the third equation that has the refractive index as a parameter. Because of the symmetry, these equations are easily memorized but, more important, useful insights can be readily gained from the relations. For example, the radius of curvature is inversely proportional to the phase velocity and to the component of the gradient of phase velocity which is normal to the ray. Thus, the ray will curve sharply when either the phase velocity or

the gradient component perpendicular to the ray becomes large. Furthermore, it can be seen that the component of the refractive index gradient along the ray has no effect on it. These insights would be difficult to glean from the sinusoidal form of Snell's law although, as will be shown, the two forms are equivalent.

Using the relation $\partial(1/\mu)/\partial\rho = -(1/\mu^2)\partial\mu/\partial\rho$, the differential form can be written with symmetry, in terms of the refractive index, although a minus sign is introduced. This has been done to point out the similarity between the following three equivalent statements of the differential form of Snell's law.

Phase Velocity	Wavelength	Refractive Index
$\frac{\partial v}{\partial \rho} = \frac{v}{\rho}$	$\frac{\partial \lambda}{\partial \rho} = \frac{\lambda}{\rho}$	$\frac{\partial \mu}{\partial \rho} = \frac{-\mu}{\rho}$

5. Equivalence of the Trigonometric and Differential Forms

The trigonometric and differential forms of Snell's law are so different in appearance that it seems worthwhile to provide a demonstration of their equivalence. To see how this can be done, we note that the trigonometric form is usually applied across a discrete boundary (having no thickness) between two homogeneous media. The differential form applies whether or not the medium is homogeneous, although the rays bend only when the refractive index varies. This suggests that we approximate the discrete boundary between homogeneous media by a smooth transition region to which we can apply the differential form. From such an application, it should be possible to derive the trigonometric form; this is the line of attack followed below.

To readily apply the differential form, it will be assumed that the transition region is constructed as sketched in Fig. 29. The phase velocity, v, changes from its initial value v_1 to its final value v_2 as a function of distance across the transition in such a manner that the radius of ray curvature, ρ, remains constant. Within the region of variation, the direction of ∇v is everywhere perpendicular to the interface.

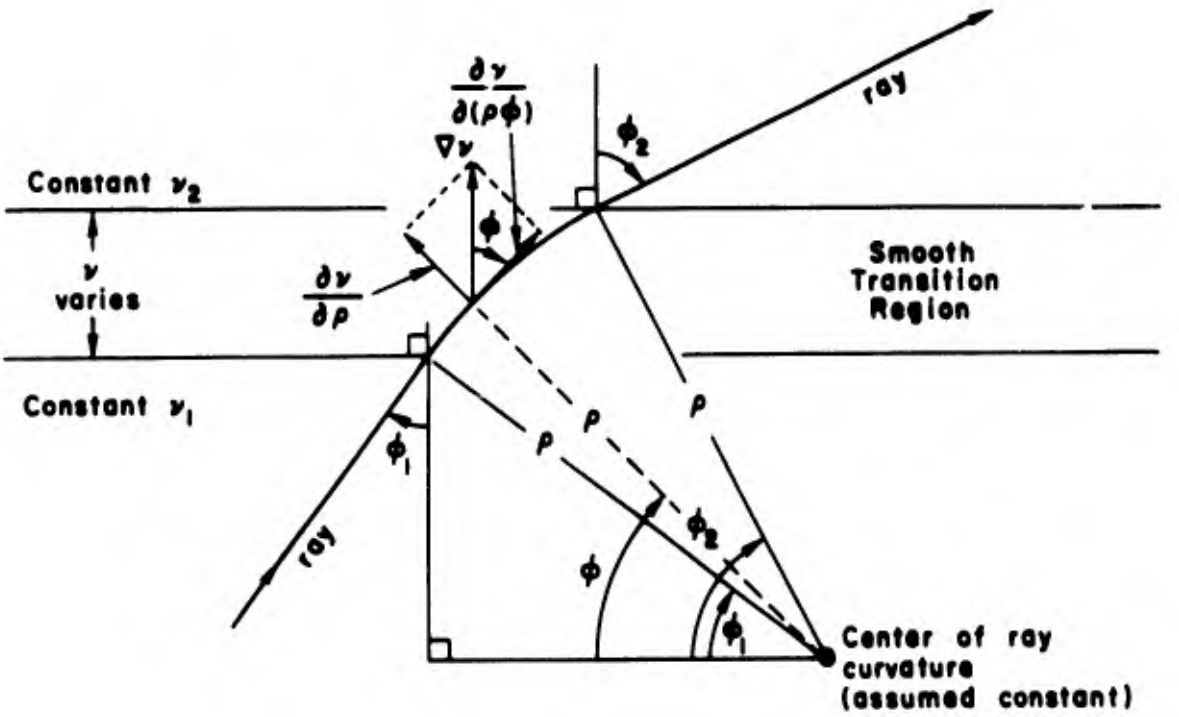


Fig. 29. RAY TRAJECTORY USED TO TEST THE EQUIVALENCE OF TWO FORMS OF SNELL'S LAW.

Since ρ is constant, the differential formula requires that $\nu/(\partial\nu/\partial\rho) = \rho$ = constant. From the vector triangles, $\partial\nu/\partial\rho = \partial\rho = \nabla\nu \sin\varphi$ and

$$\frac{\partial \nu}{\partial \varphi} = \rho \nabla\nu \cos\varphi = \rho\left(\frac{\partial\nu/\partial\rho}{\sin\varphi}\right)\cos\varphi = \nu\,\frac{\cos\varphi}{\sin\varphi}$$

Now

$$\nu = \nu_1 + \int_{\varphi_1}^{\varphi} \frac{\partial \nu}{\partial \varphi}\, d\varphi = \nu_1 + \int_{\varphi_1}^{\varphi} \nu\,\frac{\cos\varphi}{\sin\varphi}\, d\varphi$$

This is an integral equation for which the solution may be obtained from the sinusoidal form, i.e., $\nu_1/\sin\varphi_1 = \nu/\sin\varphi$. Solving for ν and substituting in the integral equation,

$$\nu = \nu_1 \frac{\sin \varphi}{\sin \varphi_1} = \nu_1 + \int_{\varphi_1}^{\varphi} \left(\nu_1 \frac{\sin \varphi}{\sin \varphi_1} \right) \frac{\cos \varphi}{\sin \varphi} \, d\varphi$$

$$= \nu_1 + \frac{\nu_1}{\sin \varphi_1} \int_{\varphi_1}^{\varphi} \cos \varphi \, d\varphi = \nu_1 + \frac{\nu_1}{\sin \varphi_1} (\sin \varphi - \sin \varphi_1)$$

Q.E.D.

Thus, application of the differential form of Snell's law leads to a homogeneous linear integral equation of the second kind whose solution is provided by the sinusoidal form. Under the restrictions of continuity of classical physics, such integral equations can have only one solution [30]; thus, the equivalence of the two forms is established. For some equations of this type, it is possible to work directly from the equation to deduce the solution. However, when we suspect that we know the solution, we may try it out, working backwards, so to speak, to see if our function is truely a solution of the integral equation. This may seem intuitively unsatisfying, rather like cheating; nevertheless, it is entirely valid and by far the most efficient way for solving such problems.

6. A Refraction Law for Sound in a Windy Atmosphere (Mark 6 Raytracing)

The following is a derivation of a refraction law for the propagation of sound in an atmosphere undergoing translation and rotation, that is, with wind. This has been checked by the author's colleagues and has been used as a basis for raytracing, but we have never seen it derived anywhere else in this or in an equivalent form. If any reader can offer an independent derivation or comment to support or challenge the accuracy of this derivation, it would be appreciated.

The sound velocity, ν, is a scalar field as before. The wind velocity has a translational component, $\bar{v}$, and it has a rotational component, ω. This calculation is in only two dimensions, so the relations between the sound and the wind are defined according to the sketch

of Fig. 30. Here, y is the component of the wind velocity parallel to the gradient of the sound velocity. For the calculation of refraction due to sound velocity changes, we will use the trigonometric form of Snell's law with an interface between two homogeneous media. Thus, x is the component of the wind velocity which is parallel to the interface and which, in turn, is necessarily perpendicular to the gradient of the sound velocity. It will be assumed that $\nu \gg |\bar{v}|$, so that $\partial\bar{v}/\partial(\text{time})$ may be neglected. To eliminate time altogether, we will again use the convention that the delay during the wavefront transit from its first to its second position is unit time, considered to be infinitesimally short.

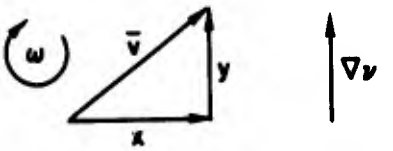


Fig. 30. VECTOR DEFINITIONS.

Figure 31 shows two positions of the wavefront at successive instants with refraction computed due to the effect of both wind and

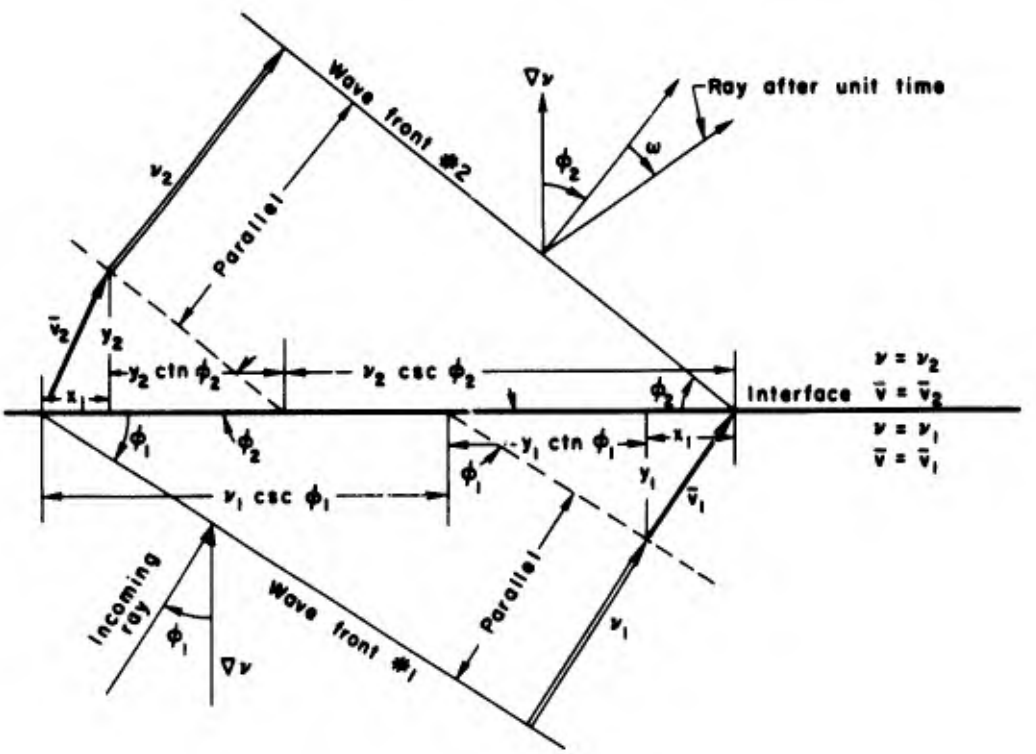


Fig. 31. GEOMETRY USED TO DERIVE SNELL'S LAW IN A WINDY ATMOSPHERE (COMPARE WITH FIG. 25).

sound. The angular wind velocity is not included because, in the computer implementation, we rotate the ray by an angle ωt after every calculation of Snell's law, where t is the transit time of the ray from the previously calculated position. Thus, ω does not enter into the refraction relation derived from Fig. 31.

Within this figure are the results of a number of simple applications of trigonometry. These serve to provide two different sets of three components for the total distance along the interface, and the sets are necessarily equal. Thus,

$$\nu_1 \csc \varphi_1 + y_1 \operatorname{ctn} \varphi_1 + x_1 = \nu_2 \csc \varphi_2 + y_2 \operatorname{ctn} \varphi_2 + x_2$$

$$\boxed{\frac{\nu_1 + y_1 \cos \varphi_1}{\sin \varphi_1} + x_1 = \frac{\nu_2 + y_2 \cos \varphi_2}{\sin \varphi_2} + x_2}$$

This reduces to Snell's law, if $\bar{v}_1 = \bar{v}_2 = 0$. The equation is easily programmed on a digital computer; an example of the output of such a calculation is given on Fig. 32.

It is recognized that the wind velocity cannot jump instantaneously from one value to a new value at the interface; this is done only to quantize the description of the variation of the wind velocity. The introduction of the third dimension of wind velocity or the second or third dimension of sound velocity would complicate this derivation to the point where a vector treatment would probably be needed. In fact, it should be noted that all the Snell's law derivations given here are simplified to one or two dimensions but that three-dimensional vector formulas corresponding to each form can be derived. For example, the differential form of Snell's law becomes

$$\frac{\mu}{\rho} = \left| \nabla \mu - \bar{T} \frac{d\mu}{ds} \right|$$

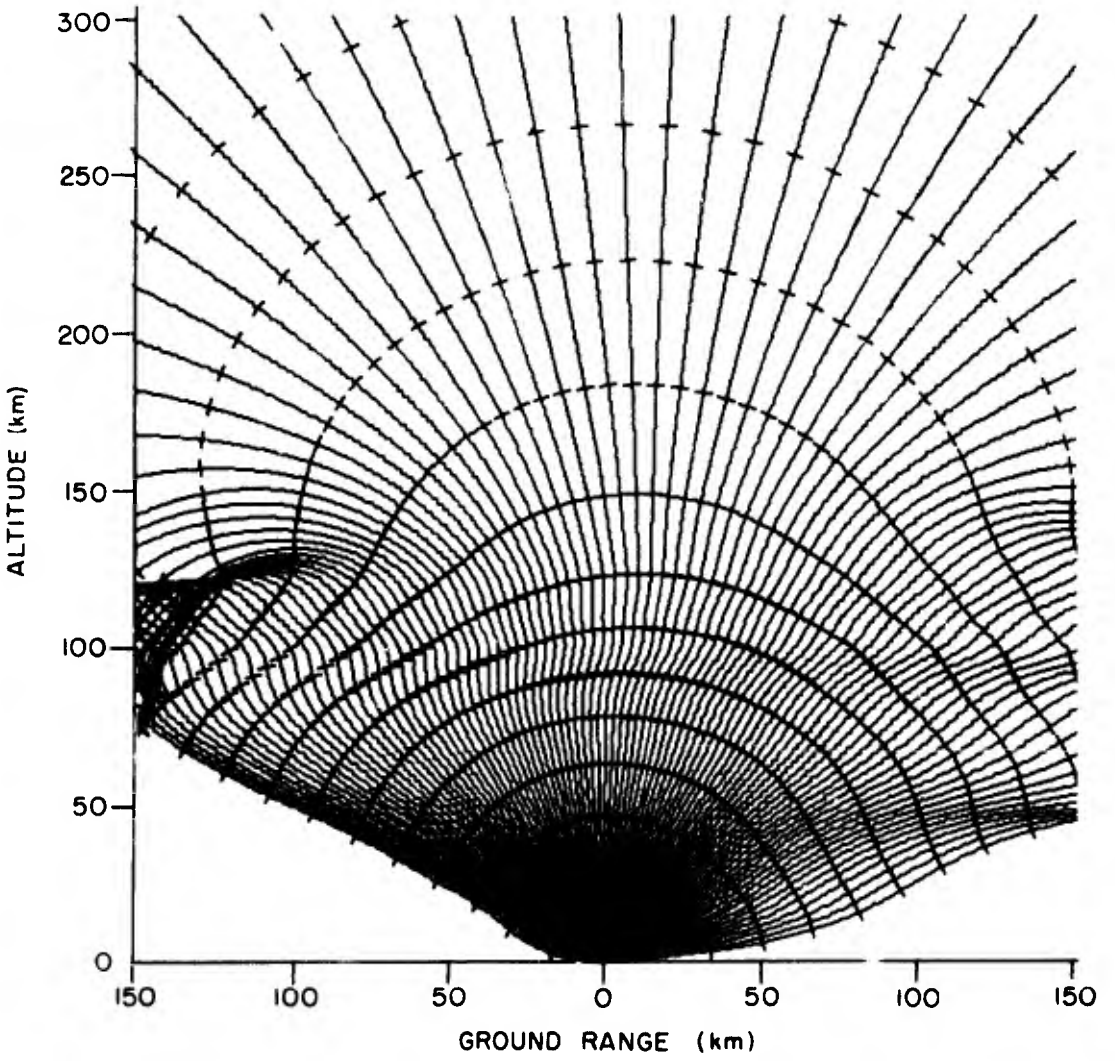


Fig. 32. ILLUSTRATION OF RAYS CALCULATED IN A WINDY ATMOSPHERE SHOWING THE LATERAL ASYMMETRY CAUSED BY THE WIND.

where $\bar{T}$ is a unit vector tangent to the ray and s is the arc length of the ray. In this report, the author has chosen to avoid such forms because many readers will be out of practice in the manipulation of vectors while still able to digest the differential form. The vector form given above is no more than a mathematical way of saying that the ray curves within the plane containing the gradient, and while in it this ray obeys the differential form given earlier.

B. Application of Bouger's Rule in Mark 3 Raytracing

The invariance of the Bouger constant along a ray permits very fast calculation; in this section, the best such method we have thus far devised is presented. The system of calculations described here serves as the basis for our Mark 3 raytracing programs. To begin, we will describe a program and present it in Fortran statements; this program has been purposely rewritten so that it can easily be comprehended by a person unfamiliar with Fortran. To simplify this explanation still further, the particular program to be derived does not contain any provision for running more than one frequency or more than one ionosphere on any given pass through the computer. Such extraneous provisions are easily devised but their inclusion in this explanation does not contribute to the discussion of raytracing.

The ionospheric model input to the program is a tabular listing of pairs of values of electron density and height. Since the ionosphere cannot be tilted under Bouger's rule, there is no need for any other parameter in the listing. Once the frequency is specified, it is possible to calculate the refractive index at each height. Then, for a ray that departs from the earth with a selected angle β_o, the Bouger constant is $r_e \cos \beta_o$, where r_e is the earth's radius. One can immediately calculate the orientation angle, β, of the ray at every listed height, h, since $\mu h \cos \beta$ also equals the Bouger constant. The raytracing problem is thus considerably reduced. We know the ray's starting location and angle, and we also know the ray angle at the selected heights. The only problem remaining is the locations of the path which best fits these data. While this may seem to be a trivial problem, it has nevertheless taken a considerable effort to try many methods in a search for the best one given here. One might do a crude job in the fitting by simply using straight lines between the listed points, but to achieve accuracy there must be a very large number of listed points; consequently, the process becomes slow and expensive. The quality of a raytracing program is measured in terms of accuracy per unit cost, since cost is usually the controlling factor which limits our capability. An economical calculation method permits the use of raytracing in applications

where more expensive methods could not be used because the priority of the application might not warrant the higher expense. To be specific, one might wish to compute a large family of rays through ionospheric models from a set of reduced ionograms. To achieve useful accuracy, a straight-line fitting calculation might cost about $50 per ionogram, whereas the method to be described here would cost about $5. (Assuming 1,000 rays per ionogram.) If there are many ionograms, this cost difference would be significant.

On the other side of the coin, it should be noted that the cruder methods will suffice for applications where only a few rays are needed. If only a total of (say) 100 rays is needed, then the relative cost per ray is negligible since the crude techniques can yield accurate answers if they are used with a sufficiently small step size. It follows that the writing of a raytracing program is easy, but the writing of a <u>good</u> one is considerably more difficult; yet it serves primarily to decrease costs. (Those who specialize in this work will recognize that I have oversimplified here, for to be sure, some techniques can never achieve the accuracy attainable by others, no matter what the step size. Nevertheless, I would like to emphasize that the better programs are superior primarily in terms of quality vs cost of operation, and that, if only a few rays are needed, a crude approach will suffice. Such comparisons cannot be applied to programs which incorporate different underlying assumptions; for example, a program that does not include the magnetic field cannot be compared easily to one that does.)

The basis of the Mark 3 programs is the fitting of second-order curves between the listed heights, having the appropriate slopes at each end of each parabolic arc. This process is very rapid by virtue of a convenient property of a parabola which is sketched on Fig. 33. Stated in words, the slope of any chord of the parabola is exactly the average of the slopes of the parabola itself at the endpoints of the chord. In the ray calculation, the "slope of the parabola" will be the ray slope obtained from known values of β, and the slope of the chord will be $\Delta h/\Delta r$. Since Δh is the difference between two listed heights, it follows that the increment of range Δr can be calculated very quickly.

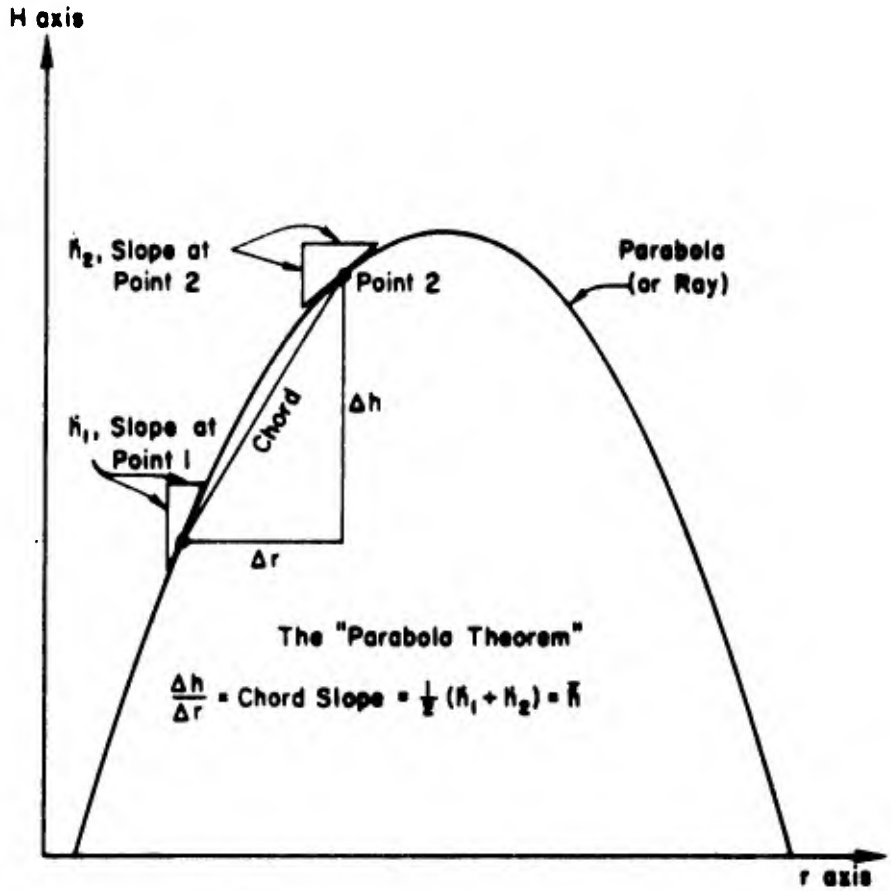


Fig. 33. THE CONVENIENT PROPERTY OF PARABOLAS.

1. The Parabola Slope Theorem

The convenient property mentioned above is not restricted to parabolas in Cartesian coordinates, and in fact, it will be applied to rays in circular coordinates. In this report, the mathematical relation will be called the "parabola theorem." Since parabolas have been analyzed for centuries, it is unlikely that this is a new discovery, but the author has not seen the property described elsewhere, so a proof of it will be given here, applicable to both Cartesian and circular coordinates. Readers who are not interested can skip directly to the next section, since this proof does not aid in raytracing if one is willing to unquestioningly accept the equation in Fig. 33.

A non-mathematical statement of the parabola theorem may clarify its significance. Consider a Cartesian coordinate system in which there is a parabola with a vertical axis. First, we select two points along the parabola and measure the slopes of the parabolic curve at the two points. Then, we calculate the average of these two slopes, that is, one-half of their sum. Next, we draw a straight line from one point to the other and measure its slope. The parabola theorem states that the slope of this new line will exactly equal the average of the slopes of the parabola at the two selected endpoints.

The parabola theorem can be carried over into circular coordinates, provided that the concept of "slope" at a point is translated into a derivative and that the slope of the line is translated to be the difference of the coordinates of the line endpoints; this will be seen in the math that follows.

Why is the parabola theorem important? The answer is quite simple. During raytracing, we know the slope of the ray at two selected heights, because we know the refractive index at those heights, and the slope can be calculated directly from Bouger's rule. Thus, we know the height difference of the points and the slope at the points, so it is an ultimately simple calculation to find the range difference of the points by means of the parabola theorem. In effect, we fit a second-order curve through the known heights, at the known angles, and that's about as good a job as can be done, since we inherently know nothing of the variation of the refractive index between the two points. (Of course, when we compute rays in an analytic ionospheric model such as a Chapman layer or a parabolic layer, then we do know the refractive index between the points. These analytic models provide us with an excellent method for testing the accuracy of the second-order fit, inherent in the application of the parabola theorem.)

The proof of the parabola theorem makes use of the symbols shown in Fig. 33. Let

$\Delta h = h_2 - h_1$ = the height of point 2 - the height of point 1

Similarly, let

$$\Delta r = r_2 - r_1$$

Define

$$\bar{\dot{h}} = \frac{1}{2}(\dot{h}_1 + \dot{h}_2)$$

where $\dot{h} = dh/dr$.

Theorem.

$$\Delta h = \bar{\dot{h}}\,\Delta r$$

Proof of Theorem.

Let

$$h = ar^2 + br + c;\quad dh = 2ardr + bdr;\quad \dot{h} = 2ar + b;\quad b = \dot{h} - 2ar$$

$$h = ar^2 + (\dot{h} - 2ar)r + c = -ar^2 + \dot{h}r + c$$

At point 1, $c = h_1 + ar_1^2 - \dot{h}_1 r_1$. Since $b = \dot{h}_1 - 2ar_1 = \dot{h}_2 - 2ar_2$, it follows that

$$a = \frac{\dot{h}_2 - \dot{h}_1}{2(r_2 - r_1)}$$

At point 2,

$$h_2 = -ar_2^2 + \dot{h}_2 r_2 + h_1 + ar_1^2 - \dot{h}_1 r_1 = a\left(r_1^2 - r_2^2\right) + h_1 + \dot{h}_2 r_2 - \dot{h}_1 r_1$$

$$h_2 - h_1 = \Delta h = \frac{\dot{h}_2 - \dot{h}_1}{2} \cdot \frac{r_1^2 - r_2^2}{r_2 - r_1} + \dot{h}_1 r_2 - \dot{h}_1 r_1 = \left(\dot{h}_1 - \dot{h}_2\right)\left(\frac{r_1 + r_2}{2}\right) + \dot{h}_2 r_2 - \dot{h}_1 r_1$$

$$\Delta h = \dot{h}_1 r_1\left(\frac{1}{2} - 1\right) + \dot{h}_2 r_2\left(-\frac{1}{2} + 1\right) + \dot{h}_1 r_2\left(\frac{1}{2}\right) + \dot{h}_2 r_1\left(-\frac{1}{2}\right)$$

$$\Delta h = \frac{1}{2}(\dot{h}_2 r_2 - \dot{h}_1 r_1 + \dot{h}_1 r_2 - \dot{h}_2 r_1) = \frac{1}{2}[\dot{h}_2(r_2 - r_1) + \dot{h}_1(r_2 - r_1)]$$

$$\Delta h = \frac{\dot{h}_2 + \dot{h}_1}{2}(r_2 - r_1) = \bar{\dot{h}}\Delta r$$

Q.E.D.

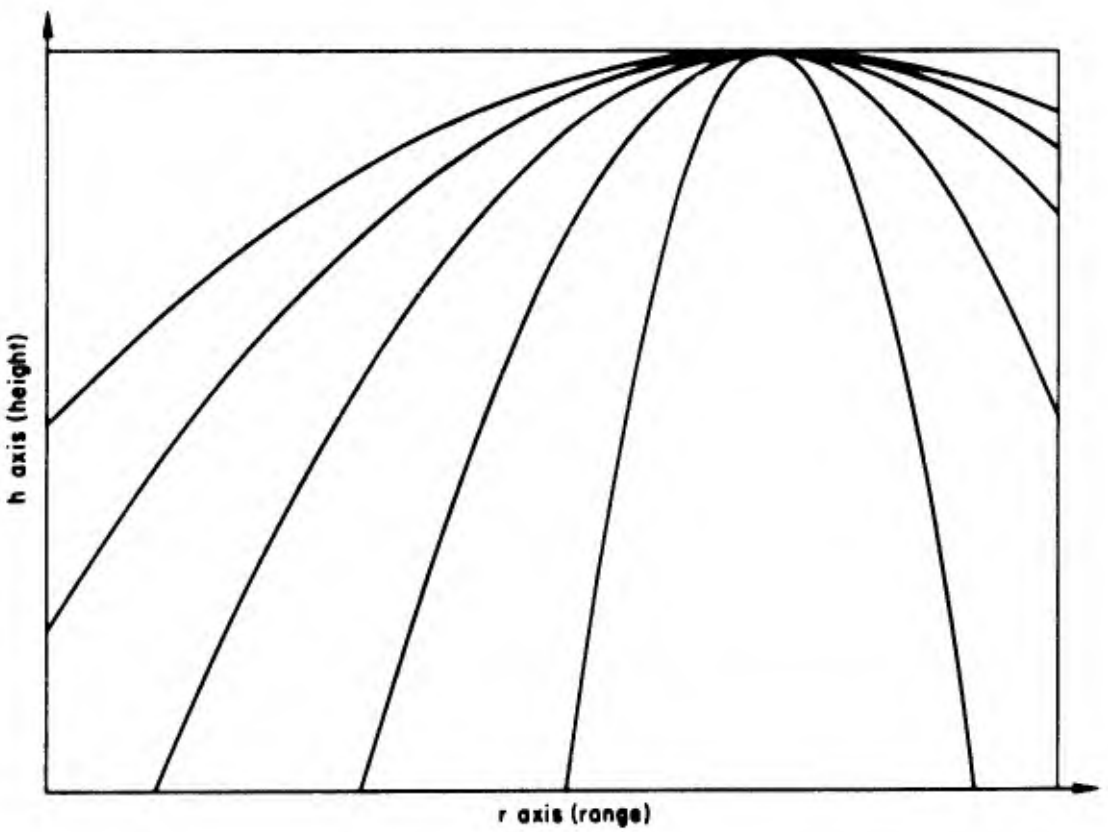
This is entirely algebraic, so it also applies in circular coordinates, if we simply substitute θ for r in the above. However, then the fitted curves are not parabolas, but the fitted sections are usually so short that the difference is immaterial. The difference between these second-order curves in Cartesian and circular coordinates is illustrated in Fig. 34.

As was mentioned previously, during the application of this theorem one knows the slopes $\dot{h}_1$ and $\dot{h}_2$ and the two heights, h_1 and h_2. Similarly, one knows the starting range, r_1, from which this calculation is the next extension. If we were working in Cartesian coordinates, we could derive the useful formula quite directly from the last line of the above proof as follows:

$$\Delta r = \frac{2\Delta h}{\dot{h}_1 + \dot{h}_2} \qquad \text{so} \qquad r_2 = r_1 + \frac{2\Delta h}{\dot{h}_1 + \dot{h}_2}$$

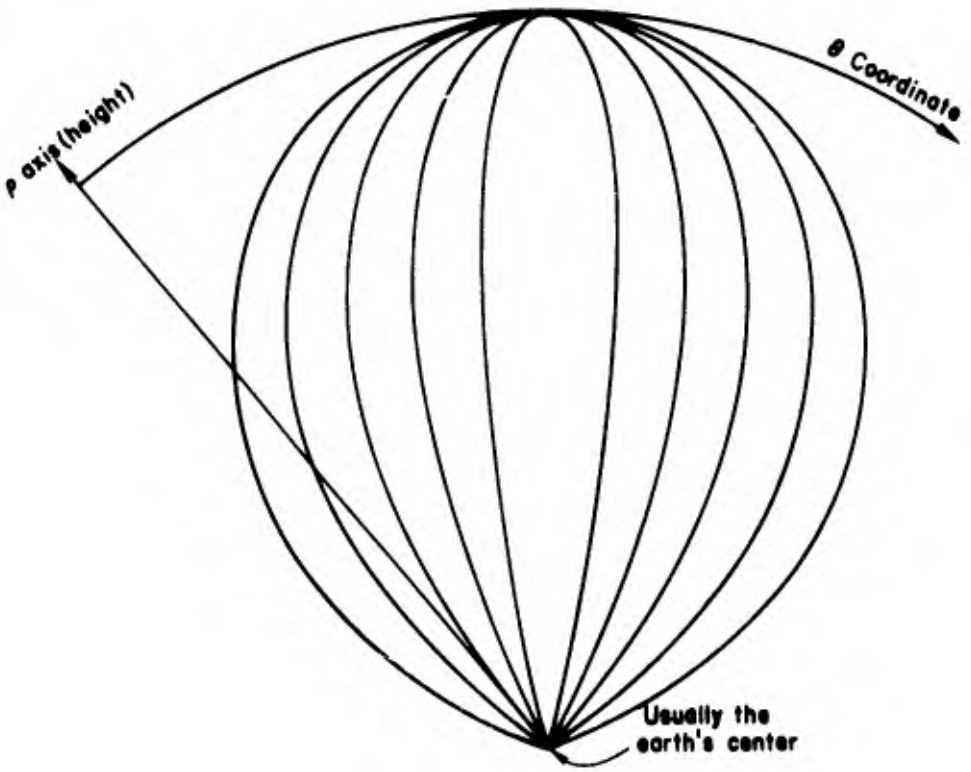
Since the author chooses to measure β, the ray vertical angle relative to the horizontal, it follows that one obtains the needed slopes through the use of the relation $\dot{h} = \tan\beta$ at points 1 and 2. The required vertical angle may be computed directly from the input table to the raytracing program which describes the ionosphere. This provides an

ordered set of matched pairs of values of height and refractive index. Given the takeoff angle of the ray at the surface and the refractive index at the surface, one immediately applies the Cartesian form of Snell's law which states that $\mu \cos \beta$ (or $\mu \sin \varphi$) is preserved everywhere along any individual ray. This provides the vertical angle at every point in the input table and one could immediately calculate $\dot{h}$ at each of the listed heights. Such a procedure could be used as the basis for a very fast raytracing program in Cartesian coordinates. We have never pursued this investigation further, since it is more realistic and only slightly more difficult to work in circular (spherical) coordinates, following the reasoning which follows.

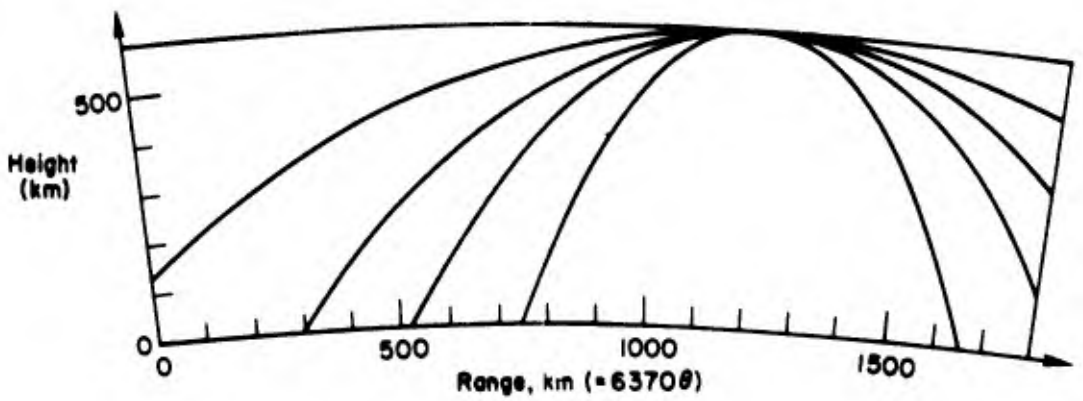


a. Parabolas in Cartesian coordinates. Notice that the curves resemble rays in a non-tilted ionosphere, that is, they bend only gradually at first, then sharply at the top (at apogee) and then again more gradually as they descend.

Fig. 34. COMPARISON OF PARABOLAS IN CARTESIAN AND CIRCULAR COORDINATES.



b. Second-order curves in circular coordinates



c. Shape of parabolas (from part b) used over the earth as compared to part (a) curves

Fig. 34. CONTINUED

As is implied in Fig. 34b, we assume circular coordinates (ρ,θ) in which there exists a parabola described by the equation $\rho = a\theta^2 + b\theta + c$. Now the algebra used to prove the parabola theorem is not restricted to the Cartesian coordinates, and this simple change of symbolism from the (h,r) to the (ρ,θ) system carries directly through the proof. It therefore follows that

$$\Delta\theta = \frac{2\Delta\rho}{\dot{\rho}_1 + \dot{\rho}_2} \qquad \text{where now} \qquad \dot{\rho} = d\rho/d\theta$$

A slight difficulty appears at this stage because it is not as easy to obtain $\dot{h}$. The reader will recall that we obtained these derivatives from the known angle of the ray which in turn was obtained from Snell's law. In circular coordinates, however,

$$\tan\beta = \frac{d\rho}{\rho d\theta} = \dot{\rho}/\rho$$

Therefore,

$$\Delta\theta = \frac{2\Delta\rho}{\rho_1 \tan\beta_1 + \rho_2 \tan\beta_2}$$

This equation is directly useful because ρ is the radius of the earth plus the height above the surface of any point. The refractive index is known at each listed height and from Bouger's rule, β can be calculated at each height. θ is the geocentric angle measured from the starting point of the ray, and its increment $\Delta\theta$ (which will later be called "A") can be calculated directly from the above equation. It is then a simple matter to find all the other needed parameters associated with the ray.

It is perhaps worth emphasizing that very little complexity is added when going from Cartesian to circular coordinates. Δh and $\Delta\rho$ are the same thing, since ρ = h + the earth's radius. Table 2 summarizes the two major differences between raytracing in Cartesian and circular coordinates.

Table 2

COMPARISON OF MAJOR FORMULAS IN CARTESIAN AND CIRCULAR COORDINATES

Formula	in Cartesian	in Circular
Snell's law	$\mu \cos \beta$ is conserved	$\rho\mu \cos \beta$ is conserved (Bouger's rule)
Parabola Theorem	$\Delta r = \frac{2\Delta h}{\tan \beta_1 + \tan \beta_2}$	$\Delta\theta = \frac{2\Delta h}{\rho_1 \tan \beta_1 + \rho_2 \tan \beta_2}$

2. Derivation of the Fortran Program

Fortran statements are so logical that little explanation of them is needed, but we do need to formulate a table relating the usual symbols to their Fortran counterparts which can be only in capital English letters. Table 3 gives these counterparts and can be used to relate the figures to the Fortran listings. We have used mnemonics wherever possible.

The program consists of logical parts. One part reads in the ionosphere and calculates the refractive index. Another starts the ray at the earth and gets it up to the base of the ionosphere. Another part operates once for each height increment of each ray inside the ionosphere, fitting the parabolic arc. (Because the last part mentioned operates so often, it is called the "main loop"; most of the ray calculation costs are incurred in this operation. Consequently, our main cost reduction strategy is aimed at speeding the main loop execution.) Still another part of the program calculates the leg of the ray which contains its apogee, that is, the top leg where the ray turns over for its return to the earth. If rays do not reflect, another part of the program processes the rays which reach the maximum height of the ionosphere. (It is said that these rays "penetrate" since they usually escape into outer space.)

Figure 35 shows the logic used to get the ray up to the base of the ionosphere. We would not want to fit a parabola section from the

Table 3

DICTIONARY OF FORTRAN TERMS FOR RAYTRACE 112E

A	geocentric angle subtended by one ray segment	ILAST	set to 1 on the last H-N card
BC	the Bouger constant, $r\mu \cos \beta$	M, N	used to indicate type of ray
BDEL	$\Delta\beta$, increment (in degrees)	PHTIM	phase time delay
BDELR	the increment (in radians)	PI2	$\pi/2$
BETA	variable β (in radians)	RANGE	distance along the earth's surface
BETADG	variable β (in degrees)	RE	radius of earth (6370 km)
BMAX	maximum β (in degrees)	RECC	reciprocal of speed of light
DSOVRC	$\Delta S/c$, free-space delay along a single ray segment	RECF	$80.6/f^2$
EN	electron density, per cm^3	RHRN	reciprocal of (height) (RN)
FREQKC	frequency, f, in kc (kHz)	RN	refractive index
GPTIM	group time delay	RNAVG	refractive index average
H	height in km, either above earth or above earth center	RNSQRD	refractive index squared
HLDOT	dh/dr on the left side	RTWLV	reciprocal of 12
HRDOT	dh/dr on the right side	SINPHI	sin (φ)
HT	apogee height above the earth	THEIID	the IID, ionosphere identification
		THETA	geocentric angle subtended

ground to the first listed height because we know the ray is straight in this region. To avoid this, we fit a straight line from the ground to height h_1 [in Fortran, H(1)], located beneath the first listed height by a distance equal to the thickness of the first shell. The first Fortran statement finds the value of the Bouger rule constant "BC" to be associated with this particular ray.

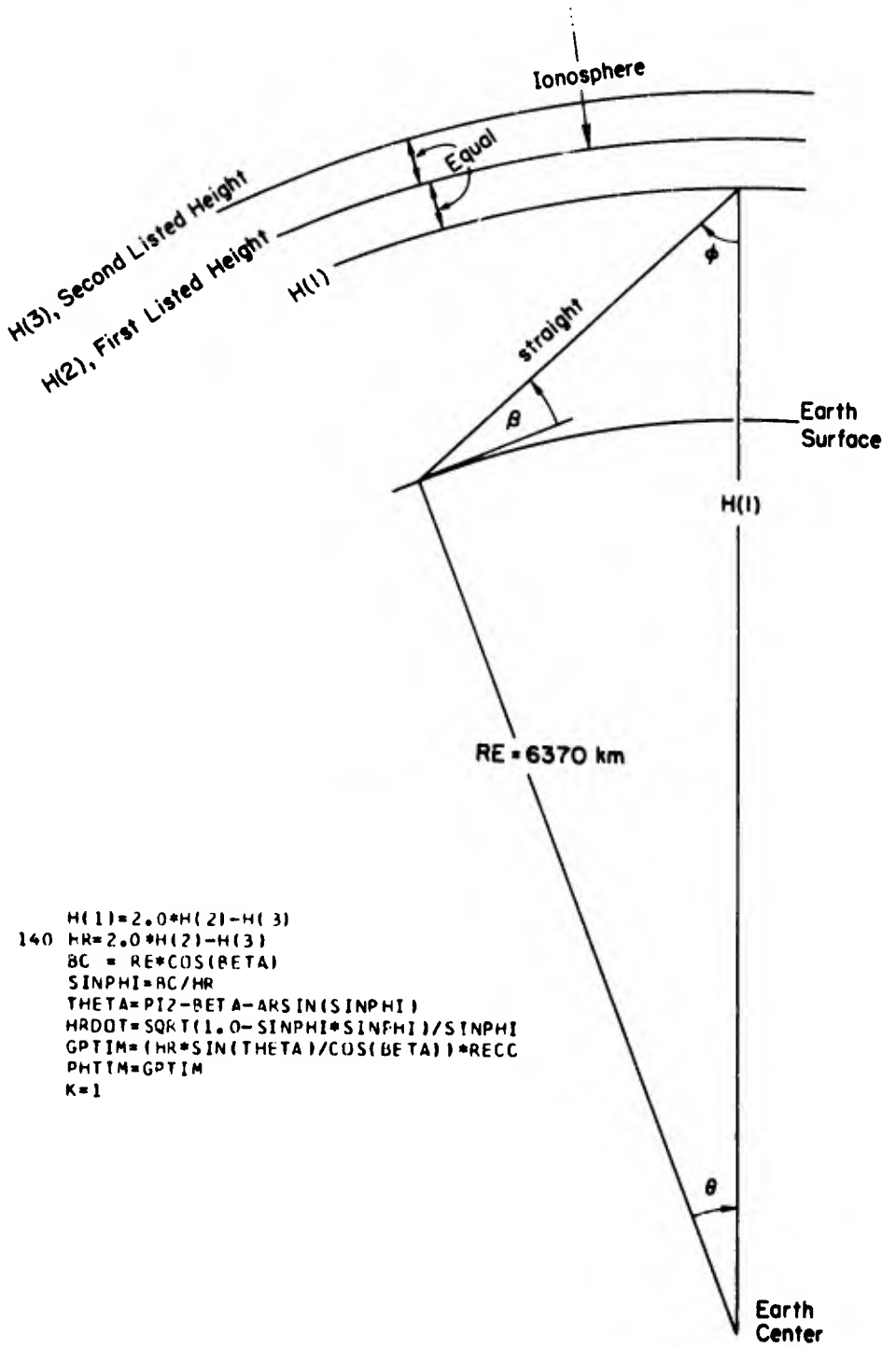


Fig. 35. STRAIGHT PORTION OF A RAY UNDER THE IONOSPHERE.

Notice that BETA is the angle of the ray measured relative to horizontal while PHI is the ray angle measured relative to the vertical. In the Fortran statements on Fig. 35, the logic is derived from the application of the law of sines to the triangle. The quantity HRDOT means the derivative of height, with respect to range (dh/r or $\dot{h}_r$), in local cartesian coordinates. Here, the r subscript refers to the "right" side of the triangle. This derivative will be ctn (φ) and is computed directly from SINPHI by a process which is fast and accurate. The parameter K is the index with which we will keep track of the layer identity as this ray proceeds upward through the shells.

Note that the first card in the electron density deck will have some height H(2) - RE, that is, the height of the first ionospheric layer above the surface of the earth. At the surface of the earth, the electron density is zero, so no card is needed in the electron density deck to describe this level.

a. The Main Loop

Figure 36 shows that part of the program most often used. For a typical ray, this sequence of statements will be operated 100 to 300 times while most other sequences are only run once. Consequently, a number of devices are incorporated to speed the operation of this "main loop."

The first Fortran statement giving SINPHI is the application of Bouger's rule. Note that RHRN is the the reciprocal of the height and the refractive index. This quantity is precalculated before rays are begun, since it is common practice to run 50 or 100 rays through a single ionosphere at each frequency and since it would be wasteful to repeatedly calculate this quantity. One could carry this practice further, if the objective is to calculate a very large number of rays in the same medium at the same frequency. For example, one could precompute and store 2.0 H(K + 1) - H(K), used in the fifth statement, and could store the squared difference and the product of H(K + 1) and H(K) for use in the seventh statement. These features are not good in all circumstances, however. They would be bad if only a few rays were needed in each medium or if computer storage were a limiting factor.

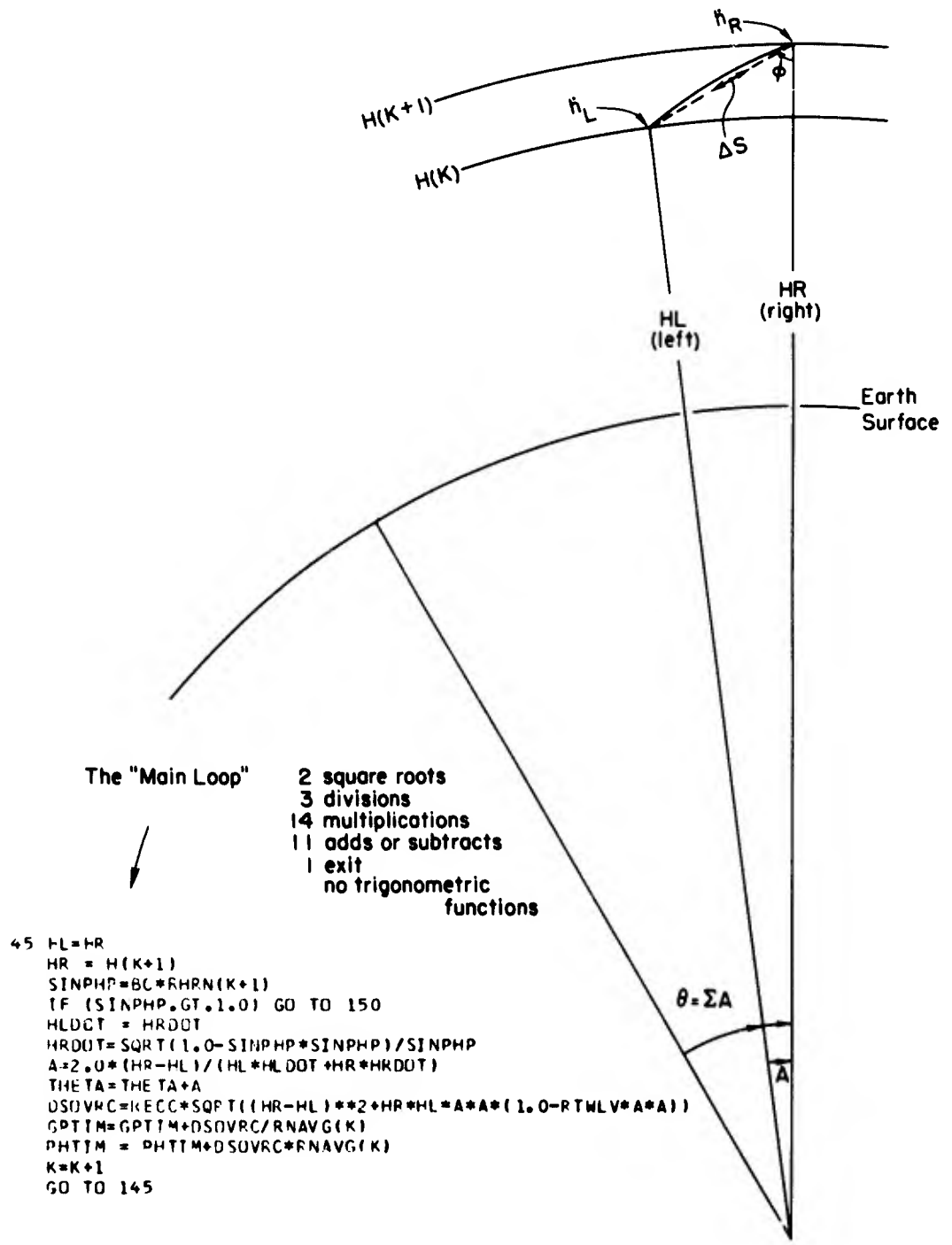


Fig. 36. GEOMETRY OF THE MAIN LOOP.

The second statement is the only way that we can exit the main loop. The SINPHI is greater than one because the ray should reflect from the level H(K + 1). Normally, we would require that there be a second possibility for exiting, when the ray reaches the maximum listed

height in the ionospheric model. However, the program execution is faster if we do not have two separate tests in the main loop; to avoid this, an artificial layer is placed above the highest listed height with a low μ that forces a ray apogee. Then, after leaving the main loop, we can detect that this false event has happened and can treat the ray properly.

The next three statements probably require no explanation. The statement defining the value of A constitutes the fitting of the parabolic segment. The simplicity of this statement is the key to the success of the method.

THETA is the accumulated range of the ray measured as an angle subtended at the earth's center. The mnemonic DSOVRC means $\Delta S/c$; similarly, RECC means $1/c$. Following this, most of the statement for the calculation of DSOVRC can be seen as a calculation of the length of ΔS through the use of the law of cosines. However, it is typical for ΔS to be about 3 km while HL and HR are roughly 7000 km; hence, the law of cosines cannot be used because the round-off error is extreme. Therefore, the series equivalent of the cosine in the law of cosines must be inserted to cancel out factors which are approximately equal to 7000^2. When this is done, we find that it is only necessary to carry terms up through A^4. The statement shown was derived in this manner. Note that RTWLV is the reciprocal of 12.

The approximation of the law of cosines is derived as

$$(\Delta S)^2 = h_R^2 + h_L^2 - 2h_R h_L \cos A$$

Now, A is very small and $\Delta S \ll h_R$ or h_L. Let $h_R = h_L + \delta$, thus δ is defined as the height increment. Then

$$(\Delta S)^2 = \underbrace{2h_L^2 + 2\delta h_L} + \delta^2 - \underbrace{2(h_L^2 + \delta h_L)}[1 - (1 - \cos A)]$$

These cancel

Expanding, we obtain

$$(1 - \cos A) = \frac{A^2}{2!} - \frac{A^4}{4!} + \frac{A^6}{6!} - \text{etc}$$

Thus,

$$(\triangle S)^2 = \delta^2 + 2h_R h_L \left(\frac{A^2}{2!} - \frac{A^4}{4!} + \frac{A^6}{6!} - \ldots \right)$$

Now, we have successfully eliminated the very large numbers causing the round-off error. The question now is: How far must we calculate the series to obtain accuracy without wasted effort? The answer, of course, depends on the standard of "accuracy." Let us choose as an example the accuracy of an IBM 7090 single precision number having 27 binary bits, equivalent to decimal 134, 217, 727. Then, we test to see how large A can be, if we use only two terms in the series. This means that we require

$$\frac{A^6}{6!} < \frac{A^2}{2(134,\ 317,\ 727)} \qquad \text{or} \qquad A < 0.04047$$

If we are dealing with an earth-based coordinate system, this is equivalent to a ground range of 258 km. We never expect to operate the main loop of an ionospheric raytracing program with a single increment that long; thus we can safely use only two terms in the series, and

$$(\triangle S)^2 = (h_R - h_L)^2 + 2h_R h_L \left(\frac{A^2}{2} - \frac{A^4}{24} \right)$$

With a few extra steps, one can readily see the source of the main loop calculation of DSOVRC.

For those readers who contemplate the use of this routine in circumstances where 2 terms are either excessive or insufficient, the following summary is offered.

Table 4

EFFECT OF APPROXIMATION LEVELS IN THE LAW OF COSINES FOR 27 BIT ACCURACY

Number of Terms	1	2	3	4
Maximum A (radians)	0.000299	0.04047	0.2[illegible]	0.584
Maximum Earth Range (km)	1.9	258 (chosen)	1465	3720

b. Time Delay in the Main Loop

The incremental group time delay for the ray segment in this shell is found by using the length of the chord, ΔS. Since this chord is actually shorter than the length of the curve, one would think that we would always have group and phase delays that are shorter than they should be. However, when this program is tested against the exact available answers, it is found that we have never detected a systematic trend for delays being shorter than they should be. Therefore, we have not found it necessary to correct for the additional length of the curved parabolic arc over that of the straight chord. In part, the problem is that the author has not found a simple (and therefore inexpensive) method of correction. If any interested reader can derive a chord length approximation of 1 or at most 2 simple terms, the author would be grateful to hear it.

Using the straight line approximation, it follows that the increment of group time is $\Delta S/c\mu$ but now the value of μ must somehow be averaged between its value at level H(K) and H(K + 1). We have tried using simple averages, the reciprocal of the average of the reciprocals, the square root of the products, and the root mean square value. The difference between the answers derived was very small and no method seemed better than others. Apparently, the kind of average to be used

depends on the nature of the variation of μ between the two listed levels. Lacking any information, it seems logical to use the simplest kind of average which is one-half of the sum. This average has been precomputed and is called RNAVG, a mnemonic for refractive index average.

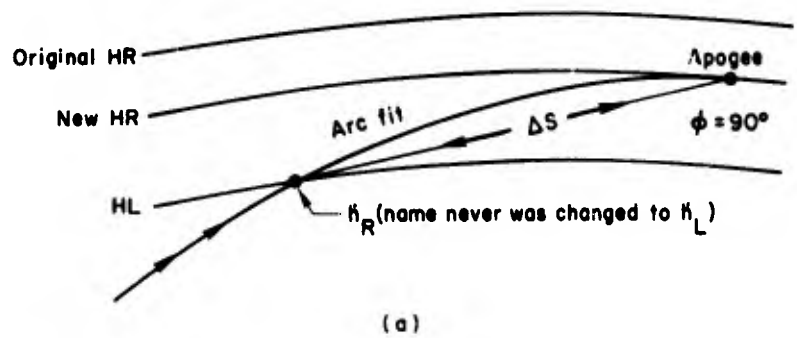
After incrementing K we return to the beginning of the main loop.

Attention is directed to Fig. 37. Upon encountering statement 150, we find that the SINPHI exceeded unity so there should have been reflection at HR. First, we test to see whether the ray is really having an apogee or if we have simply reached the maximum listed height. If the latter is the case, we will find that HR equals or exceeds an artificial height H(I + 1) in which case we go elsewhere. Assuming that this is not the case, it is necessary to find the height at which the apogee occurs.

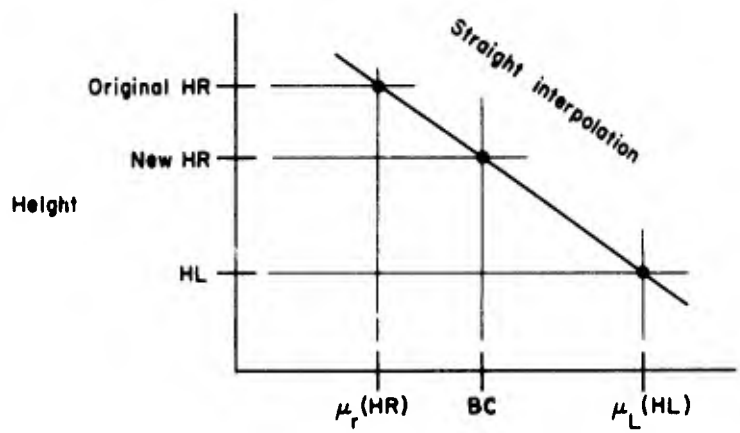
To find apogee height, we notice that in Bouger's rule the sine of the angle is unity so the product of the refractive index times the height should be equal to Bouger's constant. Furthermore, because of the ground rule that we work only from a listing of electron density vs height, we inherently know nothing about the functional variation of the product μh with respect to height, although it might be possible to assume some form of smoothness over a distance of several listed heights. We do not do this, having found that we can achieve satisfactory accuracy by using a straight interpolation which is inherently fast. This interpolation is accomplished by the fourth statement listed on Fig. 37, and it is explained by the lower part of the figure.

Here, as well as in many other places, it is possible to see where a more complicated mathematical approach might lead to better accuracy within an existing shell. However, more complicated mathematics lead to greater expense, unless we are willing to use fewer shells. Therefore, the advantage that can be achieved by a more refined approach (such as curve fitting, in this case) should be weighed against the need to use thicker shells, to achieve the same level of economy. An improvement is worthwhile only if it gives more accuracy per dollar.

```
150 IF (HR.GE.H(I)) GO TO 165
    HR=HL+(HR-HL)*(BC-RN(K)*HL)/(RN(K+1)*HR-RN(K)*HL)
    A=2.0*(HR-HL)/(HL*HROUT)
    RANGE=2.0*RE*(THETA+A)
    DSOVRC=RECC*SQRT((HR-HL)**2+HR*HL*A*A*(1.0-RTWLV*A*A))
    RNAV=RN(K)+0.75*(RN(K+1)-RN(K))*((HR-HL)/(H(K+1)-HL))
    GPTIM=2.0*(GPTIM+DSOVRC/RNAV)
    PHTIM=2.0*(PHTIM+DSOVRC*RNAV)
    HT=HR-RE
    WRITE(7,155) THETID,FREQKC,N,BETADG,BETADG,GPTIM,PHTIM,RANGE,HT,M
155 FORMAT(2X,A3,F7.0,I2,F10.4,F9.4,2F11.7,F11.4,F12.5,I2)
    BETA=BETA+BDELR
    BETADG=BETADG+BDEL
    IF(BETADG.LE.BMAX) GO TO 140
    RETURN
165 PSI=90.0-57.2957795*PHIPR
    HT=HL-RE
    RANGE=RE*THETA
    WRITE(7,155) THETID,FREQKC,N,BETADG,PSI,GPTIM,PHTIM,RANGE,HT,N
    RETURN
```



(a)



(b)

Fig. 37. THE APOGEE (a) GEOMETRY AND (b) INTERPOLATION.

The present approach is also reliable. Curve fitting is less reliable since uneducated program users might insert an electron density profile incompatible with the curve fitting idea. In summary, the simple straight interpolation, while crude, permits the use of thinner layers and it is more foolproof.

In the apogee interpolation, the factor 0.75 requires some explanation. At first, this was set at 0.5 but then the group delay was too short and the phase delay was too long by about 1 part in 10^5. Analysis of the error indicated that we were wrong in using 0.5 because it neglects ray curvature in the apogee shell. As is shown in Fig. 38, this curvature can be logically accounted for by using 0.75 instead. The resulting computed answers are then considerably improved.

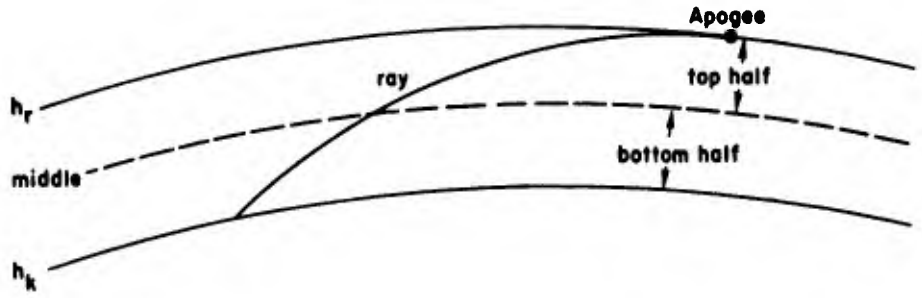
The fifth Fortran statement fits the parabolic arc. Notice that only one slope is included, because the slope at apogee is zero. The calculation of range is straightforward, and the calculation DSOVRC is identical to that in the main loop.

We are not able to use RNAVG from the precomputed array because it is the average value of μ at HL and at the original HR. What we need is the average value at HL and at the new HR which may be quite different. The calculation of RNAVG is designed to find this new average and, again, linear interpolation is used, as illustrated in Fig. 37(b).

The group time and the phase time are calculated and they, like the range, are doubled. (We stop calculating at ray apogee and double all answers, thus taking advantage of the symmetry of the ray.)

The WRITE statement, which includes format statement 155, is the standard rayset, and in this case, it includes some information that is not necessary for a non-tilted ionosphere. However, only one rayset format is used for both tilted and non-tilted ionospheres because it is easier to insert redundant information than it is to have two kinds of raysets.

Following the format statement, we increment the takeoff angle, check to see if more rays are desired, and then either start a new ray or stop the computation.



Note that most of the ray is in the top half.

Below, we approximate the ray as a circular arc in cartesian shells.

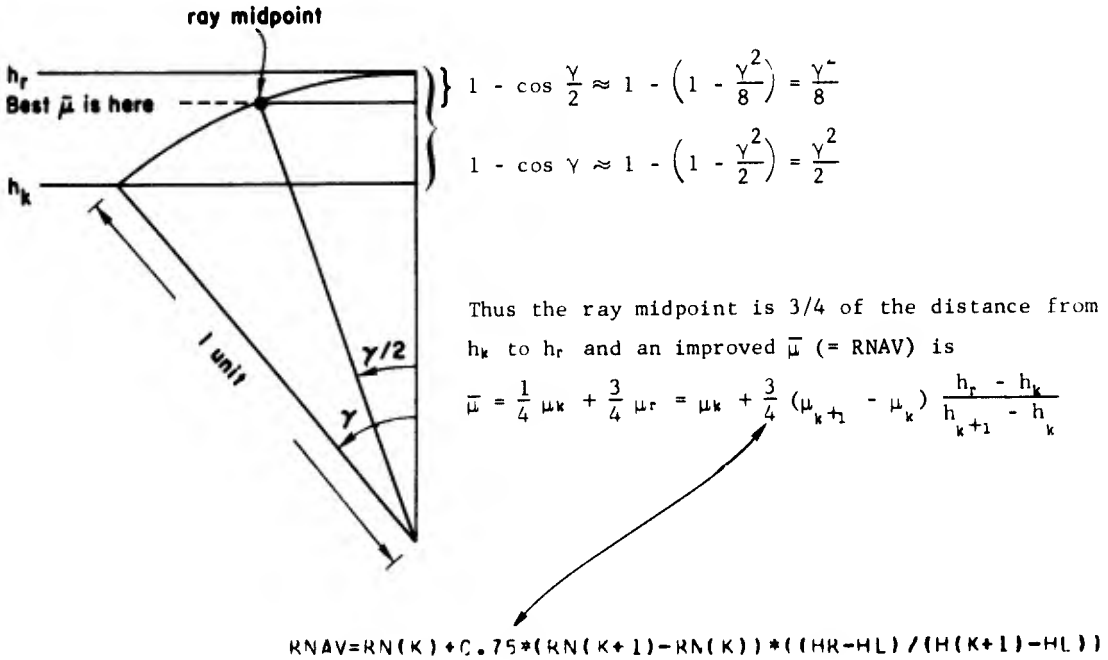


Fig. 38. DERIVATION OF THE APOGEE INTERPOLATION FACTOR.

The entire program is given on Fig. 39. In the data statement, we read the constant values to be used. Next, we read the ionosphere identification, the takeoff angle controls, and the radio frequency. Angles are changed to radians and RECF is calculated, since it will be used from here on in the calculation of the refractive index. Then the electron density profile is read and the last card, having a

```
C      RAYTRACE P85
C
       DIMENSION H(1000),EN(1000),RN(1000),RNAVG(1000),RHRN(1000)
       DATA PI2    ,    RECC    ,H(1),RE,    RTWLV    ,RN(1),M,N
      1/1.57079633,0.0033356405,2*6370.,0.083333333, 1.0 ,0,1/
  100 I=2
       READ(5,125)THEIID,BETADG,BMAX,BDEL,FREQKC
  125 FORMAT(A3,4F10.4)
       BETA=0.0174532925*BETADG
       BDELR=0.0174532925*BDEL
       RECF=80.6/(FREQKC*FREQKC)
  105 READ(5,110) H(I),EN(I),ILAST
  110 FORMAT(F8.2,E12.4,59X,I1)
       H(I)=H(I)+RE
       RNSQRD=1.0-EN(I)*RECF
       IF (RNSQRD.LE.0.0) GO TO 135
       RN(I)=SQRT(RNSQRD)
       RHRN(I)=1.0/(H(I)*RN(I))
       RNAVG(I-1)=0.5*(RN(I)+RN(I-1))
       IF (ILAST.GE.1) GO TO 115
       I=I+1
       GO TO 105
  135 H(I)=H(I-1)+(H(I)-H(I-1))*(1.0/RECF-EN(I-1))/(EN(I)-EN(I-1))
       RN(I)=1.0E-20
       RHRN(I)=1.0E+20
  115 H(I+1)=H(I)+1.0
       RHRN(I+1)=1.0E+20
       H(1)=2.0*H(2)-H(3)
  140 HR=2.0*H(2)-H(3)
       BC = RE*COS(BETA)
       SINPHI=BC/HR
       THETA=PI2-BETA-ARSIN(SINPHI)
       HRDOT=SQRT(1.0-SINPHI*SINPHI)/SINPHI
       GPTIM=(HR*SIN(THETA)/COS(BETA))*RECC
       PHTIM=GPTIM
       K=1
  145 HL=HR
       HR = H(K+1)
       SINPHP=BC*RHRN(K+1)
       IF (SINPHP.GT.1.0) GO TO 150
       PHIPR=ARSIN(SINPHP)
       HLDOT = HRDOT
       HRDOT=SQRT(1.0-SINPHP*SINPHP)/SINPHP
       A=2.0*(HR-HL)/(HL*HLDOT+HR*HRDOT)
       THETA=THETA+A
       DSOVRC=RECC*SQRT((HR-HL)**2+HR*HL*A*A*(1.0-RTWLV*A*A))
       GPTIM=GPTIM+DSOVRC/RNAVG(K)
       PHTIM = PHTIM+DSOVRC*RNAVG(K)
       K=K+1
       GO TO 145
  150 IF (HR.GE.H(I)) GO TO 165
       HR=HL+(HR-HL)*(BC-RN(K)*HL)/(RN(K+1)*HR-RN(K)*HL)
       A=2.0*(HR-HL)/(HL*HRDOT)
       RANGE=2.0*RE*(THETA+A)
       DSOVRC=RECC*SQRT((HR-HL)**2+HR*HL*A*A*(1.0-RTWLV*A*A))
       RNAV=RN(K)+0.75*(RN(K+1)-RN(K))*((HR-HL)/(H(K+1)-HL))
       GPTIM=2.0*(GPTIM+DSOVRC/RNAV)
       PHTIM=2.0*(PHTIM+DSOVRC*RNAV)
       HT=HR-RE
       WRITE(7,155) THEIID,FREQKC,N,BETADG,BETADG,GPTIM,PHTIM,RANGE,HT,M
  155 FORMAT(2X,A3,F7.0,I2,F10.4,F9.4,2F11.7,F11.4,F12.5,I2)
       BETA=BETA+BDELR
       BETADG=BETADG+BDEL
       IF(BETADG.LE.BMAX) GO TO 140
       RETURN
  165 PSI=90.0-57.2957795*PHIPR
       HT=HL-RE
       RANGE=RE*THETA
       WRITE(7,155) THEIID,FREQKC,N,BETADG,PSI,GPTIM,PHTIM,RANGE,HT,N
       RETURN
       END
```

Fig. 39. SIMPLIFIED MARK 3 RAYTRACING PROGRAM.

number 1 in the last column, signifies that the listing is finished; this is called ILAST. Since heights in the deck are measured above the surface of the earth, we must add the radius of the earth to each. Then we calculate the square of the refractive index and check to see that it is positive. After finding the index, we precompute the inverse of the product of the index with the height and also the average of two adjacent indexes, as was mentioned in the discussion on ray calculation methods.

Statement 135 is encountered, if the radio frequency is less than the critical frequency of the ionosphere. At some level, the refractive index will have a negative square, and by interpolation, we find the height at which the refractive index passed through zero. At this height, the refractive index is set equal to 10^{-20} which is essentially zero.

Statement 115 places an artificial layer above the top listed layer and sets the refractive index equal to about zero. This forces the artificial apogee which removes us from the main loop, but rays which are thus removed will also flunk the test of statement 150. These rays are then processed under statement 165 where we make a rayset for a penetrating ray and then stop computing. Statement 115 also makes an artificial layer above the processed layer by statement 135; although this layer is useless in such a circumstance, it does serve as a test object in statement 150.

By the time the operation of the program reaches statement 140, the refractive index and its various derivatives have been formed in the appropriate arrays; thus, the ray calculation proceeds according to the descriptions accompanying Figs. 35, 36, and 37. Calculation proceeds until either the maximum takeoff angle is reached or some ray reaches the maximum listed height.

In the programs actually used, there are a number of options not included here. For example, more than one frequency can be run in the same ionosphere or more than one ionosphere; if desired, penetrating ray families can be obtained; the rays can be plotted or listings can be made to provide details along the rays; tickmarks can be placed at equal increments of group or phase time delay; or, the Faraday

rotation along the ray can be computed by using a dipole field approximation and the QL formula for rotation rate.

c. Operational Practice

The program input is a tabular listing of pairs of values of electron density and height. The "main loop" of the program operates once for each height increment for each ray. Thus, the calculation is cheaper, if the height increments are large, but it is more accurate, if the height increments are small. Generally, we have found that answers are good to five or six significant figures if height increments of 1 km are used, but the best combination of cost and accuracy is achieved by tailoring the height increment so that it is small in regions where the electron density and its gradient are large. In regions where the density or the gradient is small, the increment can be made large without much loss of accuracy. In essence, it is desirable to make frequent calculations when the ray is bending sharply.

At one time it was the author's impression that the best combination of economy and accuracy could be achieved by solving Snell's law at successive increments of slant range, or of group delay, or through the use of some scheme which avoided making a large number of calculations in regions where ray bending was very slight. However, after several years of trial and error it has been my impression that the method used here is more practical. In the same ionosphere, one might want to use different height increments for rays which take off at steep angles rather than for rays which take off at shallow angles. However, it seems that the logic required to optimize increment selection requires so much programming and computation time that it is not worthwhile for the simple no-field, no-tilt case. Calculation in the main loop of this program is so fast that the expense of running it is less than or comparable with the expense needed to intelligently avoid it. One of the best features of this simple approach is that there are very few logic traps which can lead to difficulty. Therefore, it is unusually easy for someone unacquainted with raytracing to learn and to operate this program.

In other words, one of the advantages of the approach given here is that it is easy to teach to other people and that it is so simple that it seldom fails to produce a result of useful accuracy at a reasonable cost.

C. Mark 1 Raytracing, an Approximate Method for Use in Gentle Tilts

It is conventional to speak of the ionosphere as being "tilted," but consideration in greater depth often leads to the view that an ionosphere "contains horizontal gradients of electron concentration." The reason for this difference of viewpoint is probably attributable to the way in which most people learn ionospheric structure. In introductory texts, the ionosphere is usually treated as a parabolic or Chapman layer. If one introduces a horizontal gradient and then examines the lines of constant electron density on the underside of the layer, one finds that they are all tilted downward slightly in the direction of the gradient. The easiest way to explain the effect of such a gradient is by simply tilting the entire Chapman or parabolic layer. The resulting explanation is roughly correct and, while it is indeed very useful for introductory purposes, it is somewhat misleading because it conveys the idea that the real ionosphere is structured that way.

In fact, the tilt angle of the electron density contours are a function of height over any single point on the earth's surface. The isodensity contours will be vertical in those regions where the layer maxima or minima occur. Suppose, for example, that there is an E layer with a valley above it and, still higher, an F layer. Then at the nose of the E layer and at the nose of the F layer and at the bottom of the valley, the contours of constant electron density will be vertical if there is only a small horizontal gradient. This occurs because the vertical gradient goes to zero, so the horizontal gradient assumes complete control.

The considerations outlined above may seem to be of trivial importance, but they do underlie several different approximate methods for computing rays in "tilted" ionospheres. For example, some people do indeed compute rays in a non-tilted ionosphere and interpret their results in terms of an earth's surface which is off-center, thus effectively tilting the ionosphere. Another very simple method which has been used is the computation of rays in a tilted, flat slab ionosphere.

An alternative method has been to compute rays in an ionosphere which is a parabolic layer in which the maximum density changes

horizontally while all other parameters remain fixed. This ionosphere has a constant percentage change in electron concentration per unit horizontal distance; in such a model, the tilt is a very strong function of height. This approach is more in keeping with the idea of gradient rather than tilt.

The purpose of using these approximations is to achieve lower cost; for indeed, this is the only advantage. Now that digital computers are reasonably reliable, we can compute with excellent rigor the path of a ray in an arbitrarily structured ionosphere, but the cost of doing so is prohibitive for many purposes. When thousands of rays are needed but great accuracy is not needed, then some form of approximation is appropriate to bring the cost in line with the value of the results.

The Mark 1 raytracing routine is an attempt to provide an approximation level which is adequate for computing rays in an ionosphere derived from predictions or from some combination of predictions and vertical incidence soundings. These sources of ionosphere data are subject to considerable error and it would usually be inappropriate to use an exact, expensive method. On the other hand, we do, in this circumstance, have quantitative measures of ionospheric thickness, critical frequency, and horizontal gradient from the predictions. From vertical soundings, we may obtain a fairly good profile of concentration vs height. Most of this information would be lost if we simply used a parabolic layer with a tilt or a gradient. To meet this need, we want to find a technique which uses all the structural information available and yet is nearly as fast as Mark 3.

When horizontal gradients exist, one can no longer use Bouger's rule. However, note that Mark 3 raytracing could be done by solving Snell's law in the form $n_1 \sin \varphi_1 = n_2 \sin \varphi_2$ at each height along the path of each ray. This would be inefficient but would produce exactly the same result. The equivalence can be seen from the derivation of Bouger's rule given earlier, where the rule was derived from Snell's law in precisely this same manner.

First, we consider modifying Mark 3 raytracing by eliminating Bouger's rule and using the Snell's law form as described. Then, as

input data we use the refractive index which varies as a function of height; this can easily be made to be a function of distance. With this easy step, it is therefore possible to take into account the changing value of the refractive index at each new point along the ray. However, to do the ray calculation exactly, we also need to take into account the change in the gradient of the refractive index for the gradient establishes the coordinate system in which the angles are measured before applying the relation $n_1 \sin \varphi_1 = n_2 \sin \varphi_2$.

A casual inspection of the problem has led most of us to expect that its solution will be simple. However, we have tried several different lines of attack on this problem without the success of writing a program with fool-proof logic. The problems become severe because of the behavior of the gradient direction, mentioned in the preceding section. As one goes through the maximum density of the E layer with a raypath, for example the gradient will change from upward to sideways to downward in the space of a very few kilometers, and very thorny logic problems arise when one tries to derive a rigorous raytracing program from the sinusoidal form of Snell's law. Results of our best effort to date will be given in the description of Mark 4 raytracing. It is the author's conclusion that some other method would work better than the sinusoidal Snell's law; this too will be discussed.

In Mark 1, we avoid many logic problems by assuming that the gradient remains vertical throughout the raypath. This works well as long as the application of this program is restricted to ionospheres in which the horizontal gradient is small relative to the maximum vertical gradients encountered along the raypath. The reason for this is that the gradient direction varies rapidly in regions where the vertical gradient becomes very small, but in exactly the same regions, the ray does not bend very much. This can be seen from the differential form of Snell's law; this form may be written in two ways, i.e.,

$$\frac{\partial \mu}{\partial s} = -\frac{\mu}{\rho}, \qquad \text{or curvature} = \frac{1}{\rho} = -\left(\frac{1}{\mu}\right)\left(\frac{\partial \mu}{\partial \rho}\right)$$

Throughout most of the ionosphere traversed by an oblique radio ray, the refractive index varies between 1 and 0.7 or 0.8 Thus, the first term $(1/\mu)$ does not change very much along the path of the ray. The amount of curvature is therefore primarily dependent on the magnitude of the second term, the vector component of the gradient which happens to be perpendicular to the raypath. At the peaks of the layers, the gradient direction may be quite non-vertical but its size is small; therefore, the curvature of the ray is small, with the result that we can neglect the direction of gradient without incurring a major error.

Note that as a consequence of this approximation, we cannot compute rays with Mark 1 in any medium where there is a strong horizontal gradient. The program would be useless, for example, in the calculation of a raypath within a chemical release that has a strong effect on the ionosperhic electron density.

A built-in logic trap must be avoided in the application of this raytracing method. It occurs as follows: Suppose we choose to apply $\mu_1 \sin \varphi_1 = \mu_2 \sin \varphi_2$ at a succession of points along the ray, measuring φ relative to local vertical as we go. This is indeed a simple calculation to program on a digital computer. However, we must choose two spots at which we will measure the two refractive indexes, μ_1 and μ_2. As it turns out, the most apparent choice does not work well, but the less obvious choice does lead to a good computer program. The latter choice is also seen to be justified on physical grounds. These two choices are illustrated on Fig. 40.

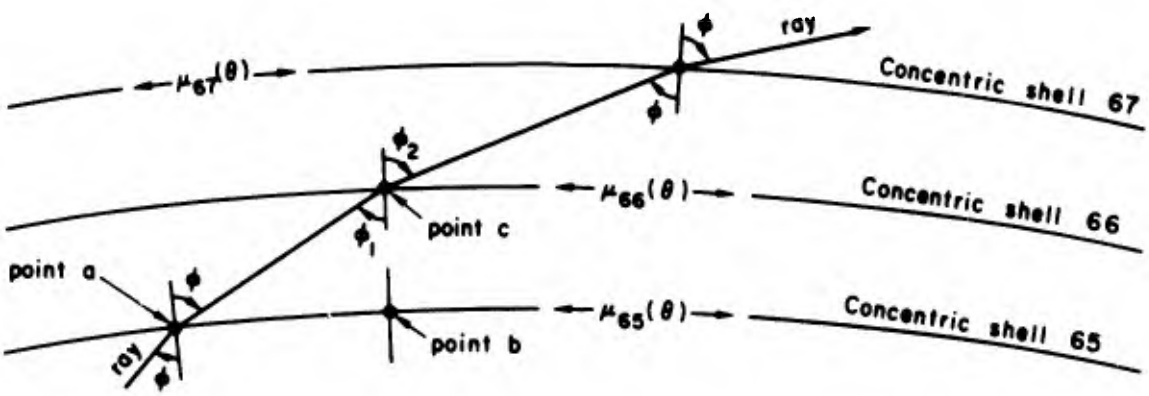


Fig 40. RAY SEGMENTS SHOWN BEFORE THE FITTING OF PARABOLIC ARCS.

This figure shows the ray as a sequence of short straight segments but, in fact, the straight lines are used only for the purpose of determining the angles ϕ. A parabolic arc will be fitted, in place of each segment, and will cause the total range to increase slightly in the illustrated example. A more precise illustration showing the geometry both before and after arc-fitting would introduce a complexity that could mask the simple point to be made here.

Consider the solution of Snell's law at point c in the middle of Fig. 40. Angles ϕ_1 and ϕ_2 are shown, and the selection of their definitions is quite straightforward. It also seems clear that μ_2 should be measured at point c because the parabolic ray segment will be fitted to have angle ϕ_2 which is intimately related to the local refractive index at point c.

However, it is not clear whether we should measure μ_1 at point a or point b. If one chooses point a, the logic trap occurs; if one chooses point b, the program works well. The difference between μ_1, as measured at points a and b, is attributable only to the horizontal gradient, since both these points are on concentric shell 65 and are obtained by evaluating $\mu_{65}(\theta)$ at the two values of θ defined by points a and b. The dependence of μ_{65} on θ actually determines the horizontal gradient in this model ionosphere.

The choice of point a would seem logical if we follow the reasoning that it represents the refractive index in the homogeneously tilted slab from which the ray segment comes. This same refractive index was μ_2 in the previous application of Snell's law at point a. If we properly account for the tilt of the gradient at point c, then we can properly use point a as the site for the calculation of μ_1. However, if we simply make the approximation that the gradient is vertical and continue to use point a, we will encounter the difficulty illustrated in Fig. 41. Figure 41a shows a particular ionospheric model by using contours of constant electron density. Notice that the gradient of electron density is perpendicular to these contours. At a height of about 120 km, the maximum density of the E region occurs, i.e., the maximum value of electron density vs height over a fixed point on the earth. At this height,

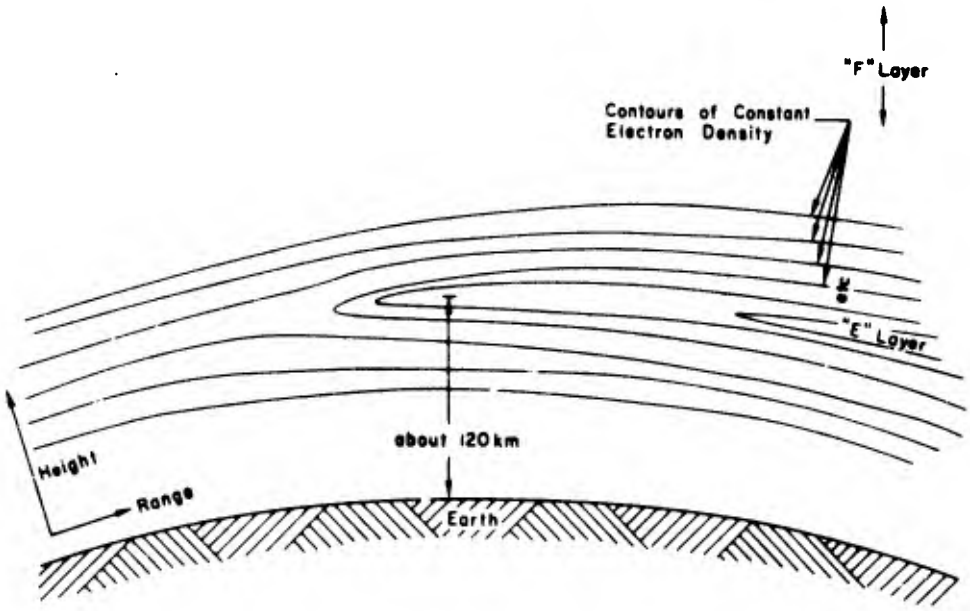
a weak horizontal gradient can be seen in the contours as an increase in electron density to the right.

Figure 41b shows a partially calculated ray traveling up through the gently tilted lower levels and arriving at 120 km where it travels nearly level. For comparison, points 1 and 2 in this figure might be considered to correspond to points a and c in Fig. 40. The fact that the ray is so nearly level means that the height difference between points 1 and 2 is very slight while the range difference is very large. Thus the change in electron density from point 1 to point 2 will be primarily due to the horizontal gradient. However, at point 2 Snell's law is solved with the angles measured relative to local vertical, and μ_2 is seen to be greater than μ_1. Thus the ray has an apogee, that is, it reflects downward at point 2. (Whenever a ray is nearly horizontal, it is not necessary to have much of a difference between μ_2 and μ_1 to cause a reflection. The interested reader can see the reason for this by evaluating the expression $\mu_1 \sin \varphi_1 = \mu_2$ for the case where φ_1 is nearly 90°.)

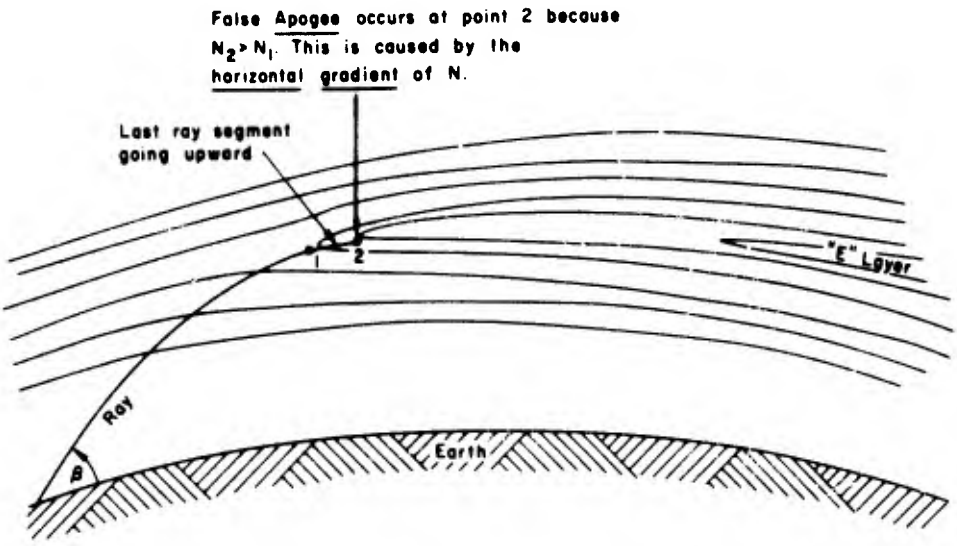
Figure 41c shows what happens immediately after the reflection at point 2. Now the ray goes downward and Snell's law can be solved for the case of a ray coming downward upon the horizontal interface between media having refractive indexes μ_3 and μ_2. Again the horizontal gradient causes μ_3 to be appreciably greater than μ_2; this causes the ray to reflect upward by exactly the same mechanism that forced it to reflect downward at point 2.

Finally, Fig. 41d shows what happens after the perigee. The ray is sent upward again but the electron density is still increasing, hence another apogee occurs. These apogees and perigees will continue in a meaningless sequence. In fact, one can see from physical reasoning that the correct ray must simply bend upward or downward in a smooth trajectory when it encounters the E layer in the situation illustrated in Fig. 41b.

The choice of refractive index μ_1 at point b will avoid these difficulties, since only the vertical gradient influences the difference between μ_1 and μ_2. This may seem unsatisfactory because it implies that the ray goes from one point to the next in a medium which is assumed to

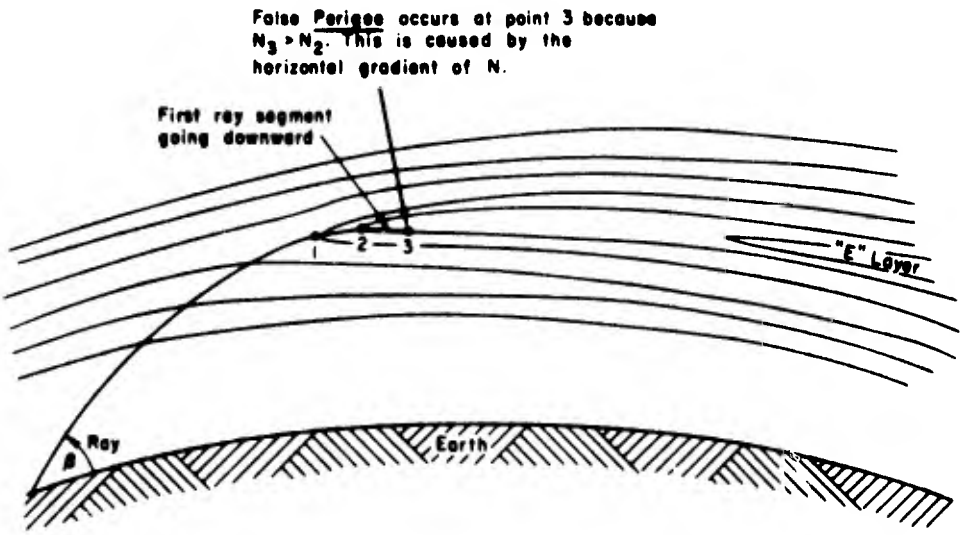


a. Representation of the ionosphere

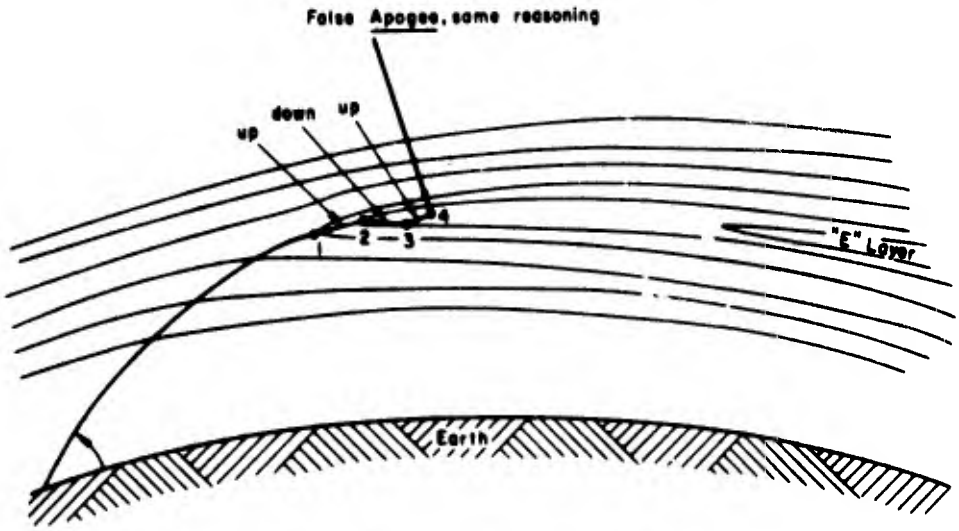


b. First false apogee

Fig. 41. ILLUSTRATION OF THE LOGIC TRAP WHICH MUST BE AVOIDED IN MARF 1 RAYTRACING.



c. First false perigee



d. Second false apogee

Fig. 41. CONTINUED

be homogeneous (because we do not solve any refraction equations during the transit) while, at the same time, the refractive index is allowed to change from its value at the first point to the next. This causes an error that, in a sense, partially cancels the error incurred in the neglect of the gradient. This is perhaps the key to the utility of Mark 1 raytracing and deserves a little further discussion here.

1. The Two Compensating Approximations in Mark 1 Raytracing

In reviewing the preceding paragraphs, the reader will find that Mark 1 raytracing follows a very simple calculation procedure through the use of two different approximations. This section will show that these two approximations cause errors of the opposite sign thus tending to cancel one another. This is very important because it explains why we are able to achieve reasonably accurate results at low cost. System tests, where end results calculated by this method have been compared against other raytracing results, appear to show the adequacy of Mark 1 for computing rays in the kind of tilts encountered in the natural ionosphere outside the auroral zone. One such test result is illustrated in Fig. 42 where two rays have been calculated in an ionosphere having a deep valley between the E and F layers. This situation was chosen because the ray becomes trapped in the valley due to the action of the horizontal gradients thus making this ray trajectory a particularly sensitive indicator of the accuracy of the tilted-ionosphere raytracing method.

As can be seen in Fig. 42, the ray takes off at a shallow angle (7.9°) and is initially almost turned back by the E layer. However, it barely makes it through the E layer on the first encounter only to be refracted downward by the F layer at a range of approximately 1000 km. Thereafter, the ray is refracted back and forth between the two layers. The fact that the rays are trapped is caused by the horizontal gradient. Furthermore, the distance between successive apogees is seen to decrease with increasing range; this is another effect of the horizontal gradient. Comparison of the two rays will show them to be nearly identical. Yet the cost ratio between the two methods is roughly 100:1, if one counts only that part of the computer program which actually calculates the

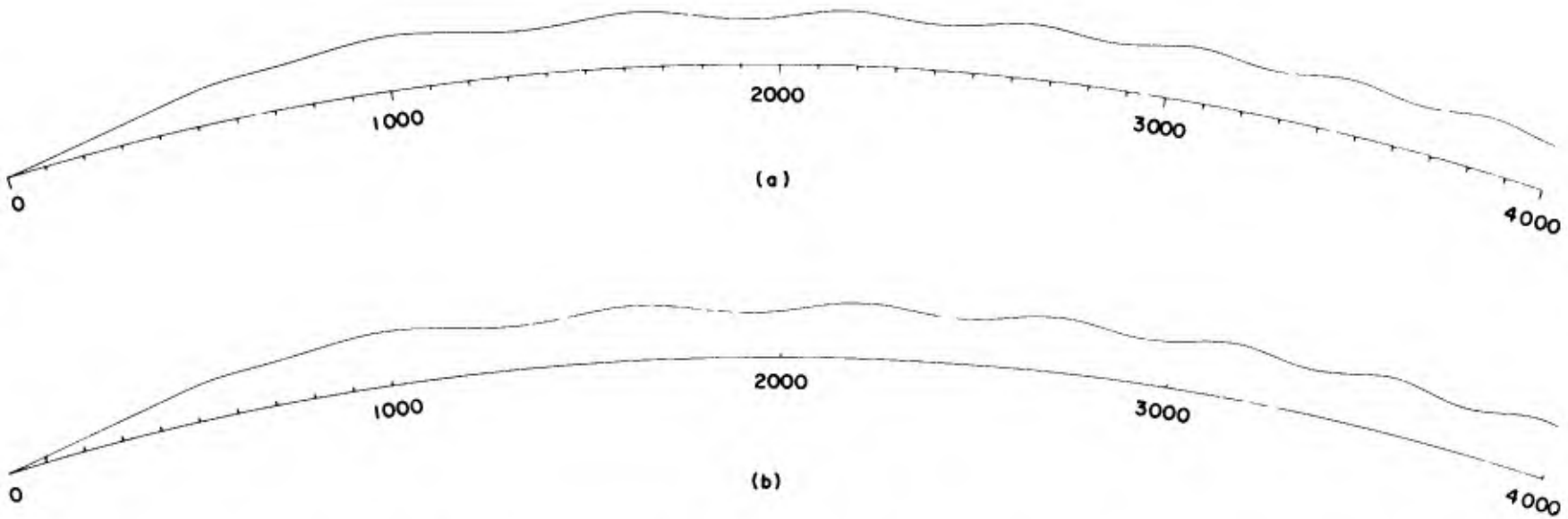


Fig. 42. COMPARISON OF MARK 1 RESULTS WITH A MORE EXACT CALCULATION. (a) Results computed by Mark 1 and (b) results computed by application of the Hazelgrove equations, Mark 8.

rays. (That part of the program which plots the rays will cost the same in each case.)

To explain clearly how the two approximations compensate for one another, it is first necessary to introduce some more explicit definitions:

> The angle of a ray measured relative to local vertical will be symbolized φ, while the angle of the same ray measured relative to the gradient direction will be symbolized ψ.
>
> The angle between the gradient and the local vertical will be called τ and, in keeping with the preceding discussion, τ will be considered small.
>
> We will follow a ray from point a to point c, as was done in Fig. 40; the intermediate point b is defined at the same range as point c and at the same height as point a.
>
> It is probably worthwhile to point out that, if the gradient of electron density (N) is in one direction, the gradient of refractive index (μ) is collinear but in the opposite direction. Thus, if $N_b > N_a$, then $\mu_b < \mu_a$, etc. To visualize the action of the ray, one thinks of it as trying to avoid regions of high electron density, that is, trying to avoid regions of low refractive index.

The ray discussions here will revolve around the drawings of Fig. 43 where three of the various conditions that can occur are shown. There are eight different cases: the ray can be going upward or downward, the vertical component of the gradient can be upward or downward, and the horizontal component of the gradient can be to the right or to the left. In this discussion, the "gradient direction" will refer to the electron density.

2. The Opposite Algebraic Signs of the Two Errors

Table 5 lists the eight possible cases which can occur. The surprising thing is that the compensation of the two approximations occurs, provided that an odd number of plus signs (+) appears in any vertical column on this table. This always occurs, thus showing that the two errors are always of opposite sign. In order to explain the derivation of this table, the first case will be considered in sufficient detail to guide the reader through the logic of the other cases.

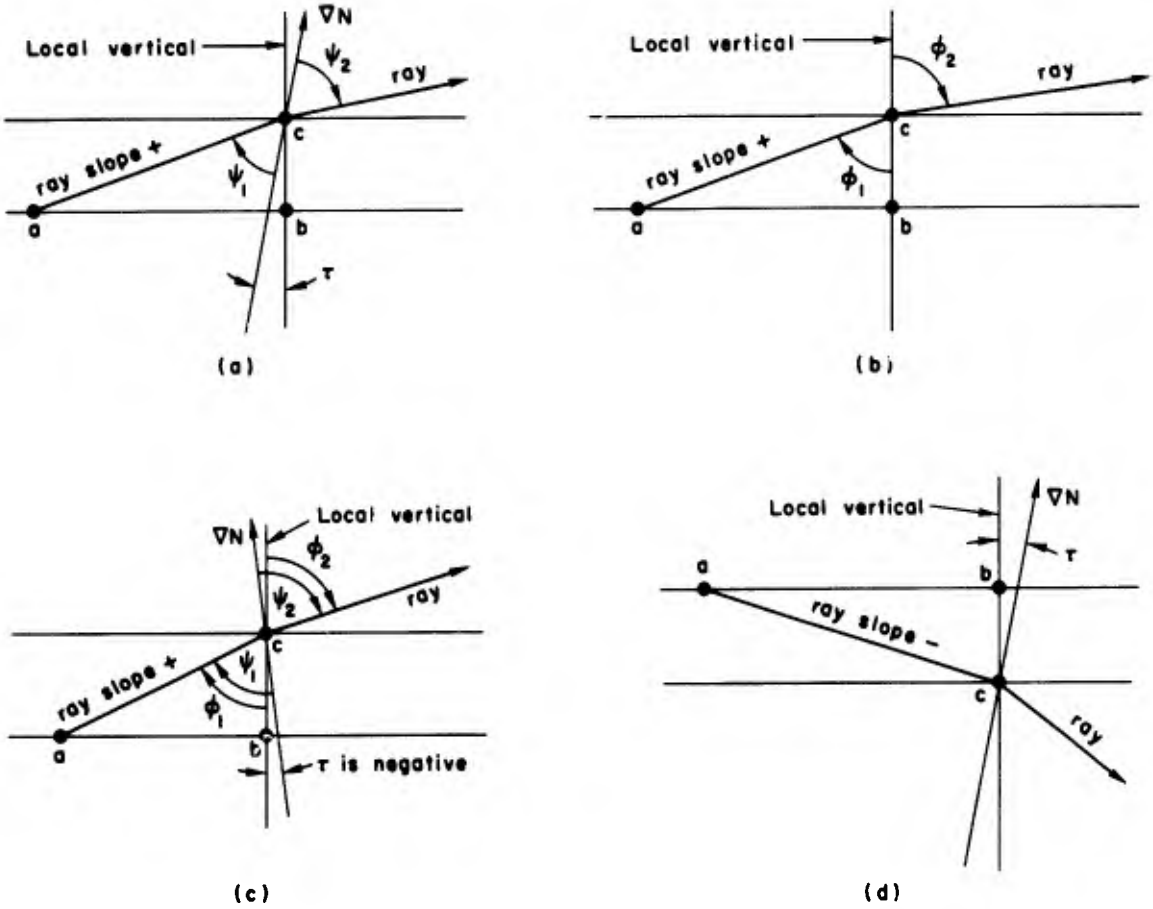


Fig. 43. GEOMETRY OF RAYS IN MARK 1 ALTERNATIVES. (a) Correct relations, Case 1, (b) Mark 1 approximations, Case 1, (c) Case 2, and (d) Case 5.

The two approximations can be defined in reference to parts (a) and (b) of Fig. 43. Part (a) shows the correct relation in which the direction of the gradient is accounted for, and the left-hand ray segment is considered to remain in a homogeneously tilted slab in which the refractive index is μ_a. The ray is considered to enter another tilted homogeneous slab above point c in which the refractive index is μ_c. Therefore, Snell's law is correctly applied through the equation $\mu_a \sin \psi_1 = \mu_c \sin \psi_2$.

First Approximation. If we simply neglect the gradient direction, that is, assume that $\tau = 0$, then we use the relation $\mu_a \sin \varphi_1 = \mu_c \sin \varphi_2$, where the angles are defined in Fig. 43b. In some

Table 5

TABULATION TO ILLUSTRATE COMPENSATION IN MARK 1 RAYTRACING*

Case Number		1	2	3	4	5	6	7	8
Gradient of N	Vertical	up	up	down	down	up	up	down	down
	Horizontal	right	left	right	left	right	left	right	left
Ray Slope		+	+	+	+	-	-	-	-
Sign of τ		+	-	-	+	+	-	-	+
Sign of $\mu_a - \mu_c$		+	+	-	-	-	-	+	+
Sign of $\mu_b - \mu_a$		-	+	-	+	-	+	-	+
Number of +'s		3	3	1	3	1	1	1	3

*Compensation occurs if the number of +'s is odd—always true.

circumstances, this approximation will cause φ_2 to be too large and, in other circumstances, it will cause the angle to be too small.

Second Approximation. Next, we assume that at point c the refractive index underneath can be measured at point b. This ignores the change in refractive index from point a to point b. We use the equation $\mu_b \sin \varphi_1 = \mu_c \sin \varphi_2'$. In some circumstances, this will cause φ_2' to be too large, and in others it will cause φ_2' to be too small.

The remainder of this section is devoted to showing that the sign of the error caused by the second approximation is always opposite to the sign of the error caused by the first. The reader who does not care to see this proven can skip the remainder of this section without loss of continuity.

Since we are interested only in the sign of the error, let us make the approximation that $\sin \varphi = \varphi$ and $\sin \psi = \psi$. Then, we obtain

the following approximate forms of the Snell's law relations before, during, and after the application of the two approximations.

Sinusoidal form	Approximate form
$\mu_a \sin \psi_1 = \mu_c \sin \psi_2$	$\mu_a \psi_1 = \mu_c \psi_2$
After the first approximation	
$\mu_a \sin \varphi_1 = \mu_c \sin \varphi_2$	$\mu_a \varphi_1 = \mu_c \varphi_2$
After the second approximation	
$\mu_b \sin \varphi_1 = \mu_c \sin \varphi_2'$	$\mu_b \varphi_1 = \mu_c \varphi_2'$

Since we are interested in the effect of the calculation at point c, let us assume that the ray leaves point a at the same orientation in all the circumstances considered here. Then, $\varphi_1 = \psi_1 + \tau$, exactly. The first approximation causes an error, ϵ_1, defined by the equation $\varphi_2 = \psi_2 + \tau + \epsilon_1$. The error introduced by the second approximation will be the difference, $\epsilon_2 = \varphi_2' - \varphi_2$. Algebraic manipulation, using these relations and the approximate forms of Snell's law, shows that

$$\epsilon_1 = \tau(\mu_a - \mu_c)\left(\frac{1}{\mu_c}\right)$$

Thus, the sign of ϵ is the sign of the product of the first two factors, since the third factor is always positive. If ϵ is positive, then φ_2 is forced to be too large by the first approximation. In the illustration of case 1, τ is positive and $\mu_a - \mu_c$ is also positive, so φ_2 is indeed too large because the tilted gradient is neglected.

Further, manipulation shows that the second approximation causes an error,

$$\epsilon_2 = \varphi_2' - \varphi_2 = \varphi_2(\mu_b - \mu_a)\frac{1}{\mu_a}$$

In this equation, φ_2 is positive in all eight cases and similarly the last factor is always positive. Therefore, the sign of ϵ_2 is the same as the sign of $\mu_b - \mu_a$. In case 1 which is being discussed, this is negative. Therefore, ϵ_2 is negative and ϵ_1 is positive, and we see that they tend to cancel.

The extension of this reasoning to the second, third, and fourth cases is quite straightforward. The reader who is interested in examining the last four cases, wherein the ray is on its downward leg, may find that the derivation is easier if the sign convention on τ is reversed so that τ is defined positive counterclockwise whenever the ray comes downward. This has the virtue that it preserves all the algebraic equations. If this logic is used, then the sign of the ray slope may be ignored, and one can simply look at the last three signs in Table 5 to see that compensation does occur, provided that an <u>even</u> number of them are positive. This viewpoint is a convenience but is not necessary. To help in visualizing the other cases, Fig. 43 shows the geometry in cases 2 and 5.

3. <u>The Relative Sizes of the Two Errors</u>

We have seen that the neglect of the gradient direction causes error ϵ_1 and that the moving of the measurement site of μ_1 from point a to point b causes error ϵ_2. It has been shown that ϵ_1 cancels ϵ_2 in the sense that they have opposite algebraic signs, but this feature would be inconsequential, if the magnitudes of the two errors were so different as to make one of them insignificant. Therefore, we will consider briefly the relative magnitude of the two errors, again using first-order approximations.

Figure 44 shows the geometrical and vector relations which will permit us to evaluate the relative sizes of the errors. Recall from the first section that, for oblique rays, the refractive index is very nearly at unity everywhere along the path. Since our interest is centered on oblique rays where the refractive index is usually above 0.7, we will, as a first step, neglect the refractive indices in the denominators of the error size expressions. These reduce to the simple form of

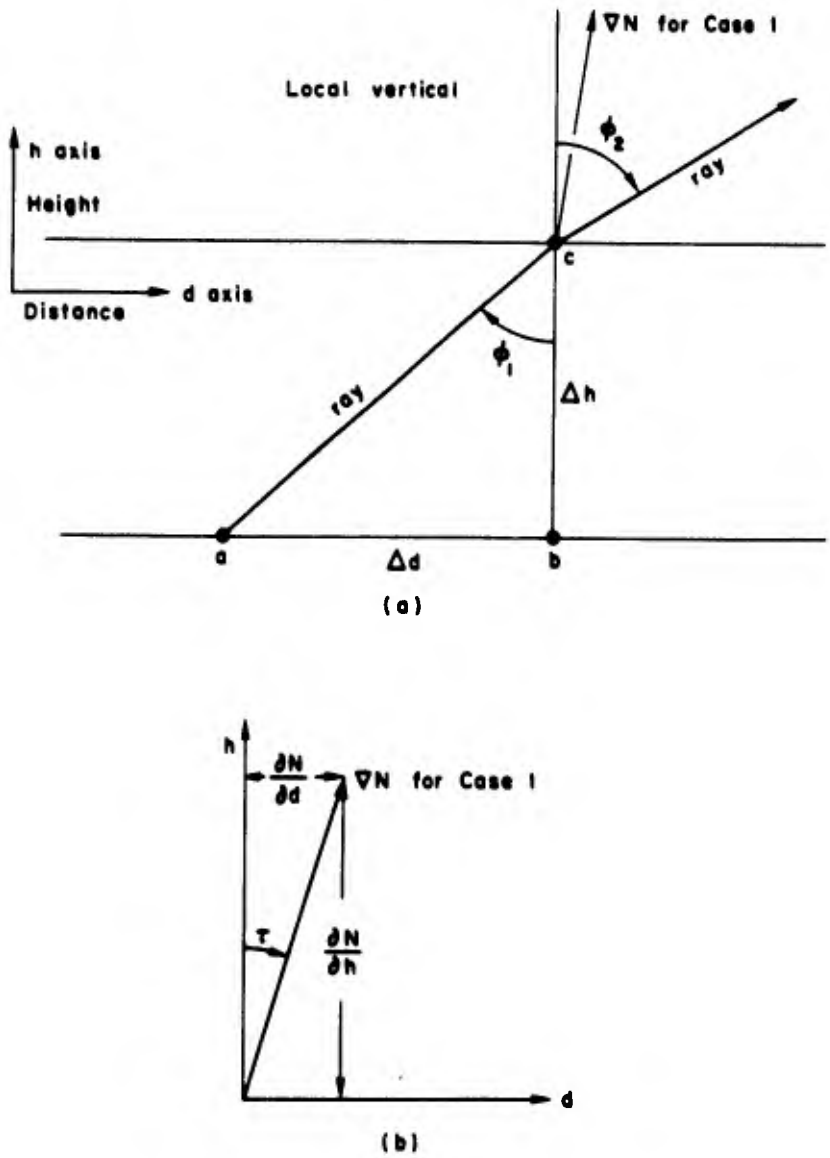


Fig. 44. GEOMETRY USED TO CALCULATE THE RELATIVE MAGNITUDES OF THE TWO ERRORS. (a) Spatial sketch and (b) vector sketch.

$$\epsilon_1 \approx \tau(\mu_a - \mu_c) \qquad \text{and} \qquad \epsilon_2 \approx \Phi_2(\mu_b - \mu_a)$$

If we can find an approximate value for the relative size ϵ_1/ϵ_2, then we can evaluate the effectiveness of the error cancellation. When the absolute magnitude is 1, the cancellation is perfect.

Working in radians and ignoring algebraic signs, the approximate magnitudes given below can be derived by inspection of Fig. 44.

$$\tau = \frac{\partial N/\partial d}{\partial N/\partial h} = \frac{\partial \mu/\partial d}{\partial \mu/\partial h}$$

$$\varphi_2 \approx \varphi_1 \approx \Delta d/\Delta h$$

$$\mu_a - \mu_c = (\mu_a - \mu_b) + (\mu_b - \mu_c) \approx \mu_b - \mu_c = (\Delta h)(\partial \mu/\partial h)$$

$$\mu_b - \mu_a = (\Delta d)(\partial \mu/\partial d)$$

With these values inserted in the expression for ϵ_1/ϵ_2, almost everything cancels out leaving the relation

$$\epsilon_1/\epsilon_2 \approx (\Delta h/\Delta d)^2$$

From this relation it can be seen that the ratio is very nearly unity, if the ray is traveling at a near 45° orientation, for then $\Delta h = \Delta d$ and the errors are of equal magnitude in opposite signs. When the ray is traveling more nearly vertically, then Δd becomes smaller and the effect of ϵ_1 is greater than that of ϵ_2. In other words, when the ray is traveling in a direction steeper than 45°, the error caused by neglecting the gradient dircction is larger than the compensating error which arises from making measurements at point b instead of point a.

When the ray is traveling more nearly horizontally, the above approximation would appenr to indicate that Δd gets very large and therefore ϵ_2 predominates. However, in the derivation we neglected a term in the expression for $\mu_a - \mu_c$. To be more proper, we should use

$$\mu_a - \mu_c = (\Delta d)(\partial \mu/\partial d) + (\Delta h)(\partial \mu/\partial h)$$

In this case, the cancellations are not so complete and the expression for the ratio is

$$\epsilon_1/\epsilon_2 = \left(\frac{\Delta h}{\Delta d}\right)^2 + \tau\left(\frac{\Delta h}{\Delta d}\right)$$

The extra term shows that indeed there is some added compensation at the lower ray elevation angles. Perhaps it is most important to notice that, when the ray is traveling nearly level, the horizontal gradient has almost no effect on it. Earlier in this section, it was pointed out that the curvature of a ray is primarily influenced by the component of the gradient which is perpendicular to the ray. When the ray is exactly horizontal, there is no component of the horizontal gradient that is perpendicular to the ray. When the ray is nearly horizontal, the horizontal gradient has some effect, but the component perpendicular to the ray is very small, so the effect is small. These considerations appear to indicate that the compensation of the two errors breaks down most badly in the region where it is least needed, that is, where the ray is nearly horizontal so that $\Delta h/\Delta d$ is very small.

D. Mark 5 Raytracing, a Long-Range Extension of Mark 1

For reasons that are primarily concerned with practical problems arising from the use of present day digital computers, we have at times found it convenient to compute long-range rays by a multiple-stage process. If a ray extends more than 5000 or 10,000 km, then there must be a great deal of ionospheric structural information placed in the computer at one time. Furthermore, as is usually the case, one does not wish to have a plot of the rays for the entire distance, only for a small selected range interval near the end of the path. If the calculation is done in several stages, then one is free to plot or not to plot the rays in each stage. Furthermore, the ionospheric coding problem is much easier because it is necessary only to code one stage at a time. For example, one might compute the rays in 5000 km increments, doing the first 5000 km one day, the next 5000 km the next day, and so forth.

There is another great advantage to this technique. Often, one wishes to calculate rays along a path and later recalculate the rays along the same path, changing only the ionospheric structure at the end of the path. It would then be wasteful to recalculate the rays along the first

portion of the path where the ionosphere did not change. When the rays are calculated in multiple stages, this waste is easily avoided. The rays are calculated over the first part of the path, and their endpoints are stored. Then the stored information can be used to start up the same rays in the second-stage calculation, using different terminal ionospheres on different occasions.

It turns out that we have already invented the program necessary to calculate the first stage. The Mark 1 raytracing program will carry the rays as far as we like and, at the termination distance, Mark 1 programs will make raysets telling the height, orientation, time delay, hop number, and frequency of each ray. This is all that we need to know, if we wish to continue calculating the family of rays. Therefore, we have modified the Mark 1 raytracing program so that it reads raysets as input data while it makes rayset output data. This is called Mark 5; it is a minor modification as far as the computer program is concerned but, to the user of the program, it is a major change. Therefore, we have given it a different Mark number since the Mark number primarily serves to guide the user.

For a time, this laboratory was actively studying round-the-world (RTW) HF signals; this study provided a good opportunity for profitable application of the Mark 5 technique. We chose to compute rays to about 500 km on each day of operation. On each day, the data were inspected. If this examination of the results indicated that our ionospheric model was inappropriate, we could go back and change only a portion and then recalculate the rays from the place where we first encountered the ionosphere. In this way, the expense was greatly reduced. Also, it was not necessary to plot the rays for the entire 40,000 km path, since we were interested primarily in the ray distribution as the rays came back to the transmitter at the end of their journey. Thus, the cost of ray plotting was reduced.

E. Mark 4 Raytracing, Using Snell's Law in Steeply Tilted Ionospheres

At the time of writing, this program has been written, but we have not completed the inevitable process of finding and removing all the logic faults. Therefore, the following description will be made quite

concise in order to keep the text as brief as possible. The operational equations which will be given are thought to be correct. At least, we have executed the program with enough success to uncover typographical and logical errors which existed in the original written version.

The technique is much like Mark 3 raytracing in the sense that we first solve for a raypath using a straightline extension from a more recently calculated position and orientation of the ray. The gradient direction is calculated and then the sinusoidal form of Snell's law is applied with reference to the local gradient. Finally, a second-order curve is fitted, that is, the parabola theorem is used here too. Now, however, the parabola theorem is applied with respect to a circular coordinate system whose center coincides with the local center of curvature of the refractive index distribution. It is this last step which leads to most of the existing complication. Nevertheless, it will be seen that the main loop appears to be reasonably short. Thus, it is expected that this program will be relatively inexpensive, if we can make it foolproof.

Before proceeding with this explanation, it is perhaps worth restating that a very promising alternate approach is available. The interested reader should refer back to Section V.A.3 entitled "Snell's Law in a Tilted Ionosphere," to find a description of the other method. To the author's knowledge, the alternate method has never been used outside Russia. Therefore, we have no basis for judging the relative merits of that approach and the one described hereafter, but the author is of the opinion that the Russian technique shows great promise. This is only an opinion, however, and could only be tested by comparing results calculated on the same computer by that method and by the Mark 4 method which follows. It may turn out that some other method (e.g., the Hazelgrove equations) is inherently superior. In short, the best method has not yet been established; it may be the Mark 4, but then again it may not be.

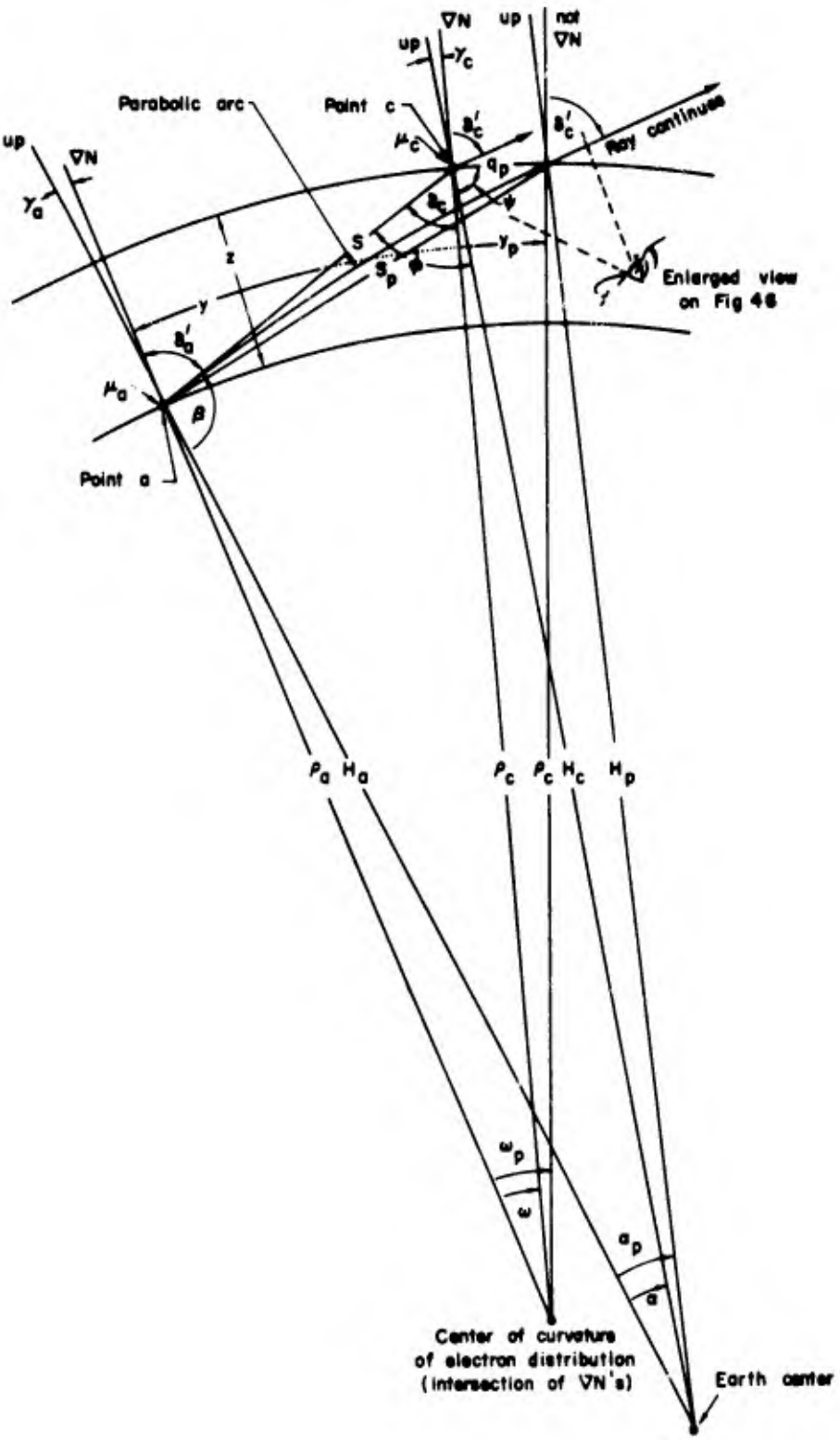


Fig. 45. GEOMETRY OF A RAY SEGMENT IN THE MAIN LOOP.

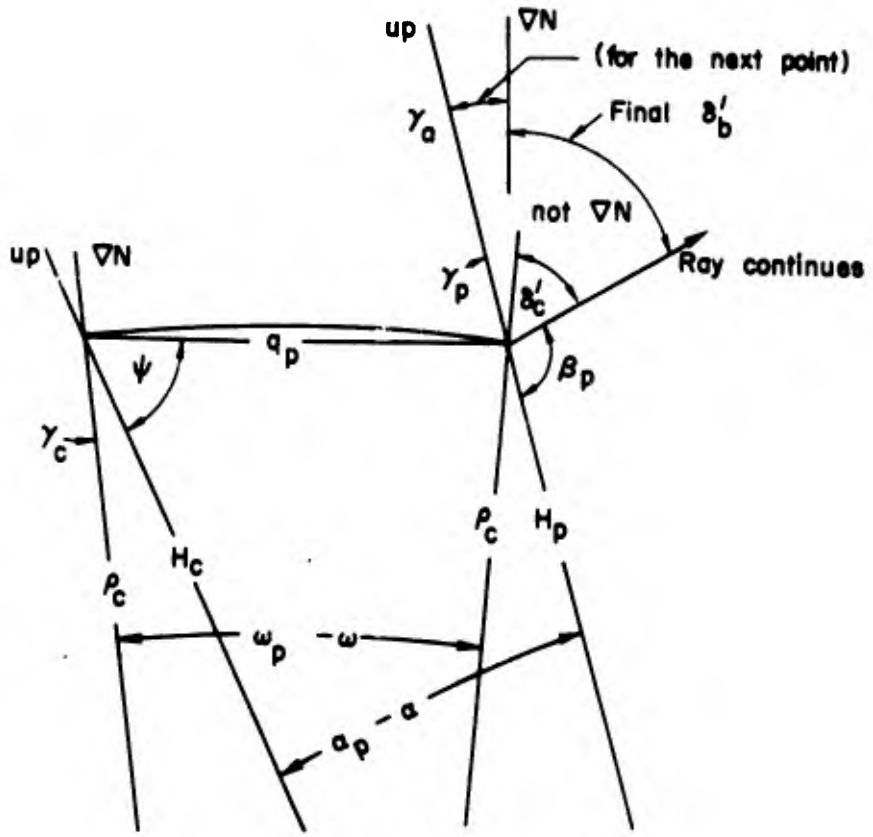


Fig. 46. ENLARGED VIEW OF A PORTION OF THE PRECEDING FIGURE.

1. The Main Loop

The main loop of Mark 4 is given by the following sequence of logic statements which relate the parameters defined on Figs. 45 and 46. The equations and logic statements are numbered for ready reference in the text which follows

Main Loop Begins Here at Point a
We know H_a, S, β, γ_a, N_a, μ_a, δ_a, T_g, T_φ, and θ (4.1)

$$H_c = \sqrt{H_a^2 + S^2 - 2H_a S \cos \beta} \tag{4.2}$$

$$\alpha = \sin^{-1}\left(\frac{S}{H_c} \sin \beta\right) \tag{4.3}$$

$$\varphi = \pi - \alpha - \beta \tag{4.4}$$

Read N_c and γ_c at H_c (4.5)

Check $|\gamma_a - \gamma_c| < 5°$ (4.6)

If $N_c > f^2/80.6$, set $\mu_c = 10^{-20}$ and skip the next step. Find new values for H_c, φ, α, and S: (4.7)

$$H_c = H_a + (H_c - H_a)\,(f^2/80.6 - N_a)/(N_c - N_a) \tag{4.8}$$

$$\varphi = \sin^{-1}\left(\frac{H_a}{H_c}\sin\beta\right) \tag{4.9}$$

$$\alpha = \pi - \varphi - \beta \tag{4.10}$$

$$S = H_a \frac{\sin\alpha}{\sin\varphi} \tag{4.11}$$

$$\mu_c = \sqrt{1 - \frac{80.6\ N_c}{f^2}} \tag{4.12}$$

$$\delta_c = \varphi - \gamma_c \tag{4.13}$$

$$\omega = \pi - \delta_c - \beta - \gamma_a \tag{4.14}$$

$$\rho_b = S \frac{\sin\delta_c}{\sin\omega} \tag{4.15}$$

$$\rho_c = S \frac{\sin(\beta + \gamma_a)}{\sin \omega} \tag{4.16}$$

$$y = \frac{1}{2}\omega(\rho_a + \rho_c) \tag{4.17}$$

$$z = \rho_c - \rho_a \tag{4.18}$$

$$\dot{\rho}_a = \operatorname{ctn} \delta'_a \tag{4.19}$$

$$\mu_a \sin \delta_c = \mu_c \sin \delta'_c \quad \text{(Snell's law)} \tag{4.20}$$

$$\dot{\rho}_c = \operatorname{ctn} \delta'_c = \frac{\sqrt{1 - \sin^2 \delta'_c}}{\sin \delta'_c} \tag{4.21}$$

$$y_p = \frac{2z}{\dot{\rho}_a + \dot{\rho}_c} \tag{4.22}$$

$$\omega_p = y_p \omega / y = \frac{2(\rho_c - \rho_a)}{\rho_a \dot{\rho}_a + \rho_c \dot{\rho}_c} \qquad \text{(Parabola Theorem)} \tag{4.23}$$

$$S_p = \sqrt{\rho_a^2 + \rho_c^2 - 2\rho_a \rho_c \cos \omega_p} \approx \sqrt{(\rho_c - \rho_a)^2 + \rho_c \rho_a \omega_p^2 \left(1 - \omega_p^2/12\right)} \tag{4.24}$$

$$q_p = 2\rho_c \sin\left(\frac{\omega_p - \omega}{2}\right) \tag{4.25}$$

$$\psi = \frac{1}{2}(\pi - \omega_p + \omega - 2\gamma_c) \tag{4.26}$$

$$H_p = \sqrt{H_c^2 + q_p^2 - 2H_c q_p \cos \psi} \tag{4.27}$$

$$\alpha_p = \alpha + \sin^{-1}\left(\frac{\omega_p}{H_p}\sin\psi\right) = \Delta\theta \tag{4.28}$$

$$\beta_p = \pi - \delta'_c - \gamma_c - \omega_p + \omega + \alpha_p - \alpha \tag{4.29}$$

Read next N_a and γ_a at (α_p, H_p) (4.30)

$$\mu_a = \sqrt{1 - \frac{80.6\ N_a}{f^2}} \tag{4.31}$$

$$\bar{\mu} = 1/2\ (\mu_a + \text{previous } \mu_a) \tag{4.32}$$

use $(S_p, \bar{\mu})$ to find T_g and T_φ (4.33)

$$\delta'_a = \pi - \beta_p - \gamma_a \tag{4.34}$$

$$H_a = H_p \quad \text{and} \quad \beta = \beta_p \tag{4.35}$$

Begin Main Loop Again (4.36)

Statement (4.1) establishes the starting point for entering the main loop of this calculation. (To understand what is meant by "main loop," refer to the description of Mark 3 raytracing.) It is assumed that the position and orientation of the ray are known at point a on Fig. 45, since that is the point at which the solution is previously calculated.

Statement (4.2) is an application of the law of cosines to the oblique triangle having sides H_a, H_c, and S.

Statement (4.3) is an application of the law of sines.

Statement (4.4) follows from the fact that the sum of the interior angles of a triangle is π radians.

Statement (4.5) in the computer program entails deciphering the coded ionospheric model to determine the electron density and the angle of the gradient of electron density at point c.

Statement (4.6) is an arbitrary test which checks to see that the gradient direction does not change very much in a single computed step. If the data did not pass this checkpoint, then we would have to go back to the beginning of the main loop and try again with a shorter step, that is, with a smaller value of S.

When rays travel nearly vertical in a nontilted ionosphere, it may be that a single step is long enough to drive the ray up past the critical region into an area where the energy cannot propagate at the radio frequency in question. In some highly tilted ionospheres which this program may encounter, it may be that a ray will be driven into such a region even though the ray is oblique. In any case, the data will not pass the test given in step (4.7) for then the electron density will exceed the critical value, $f^2/80.6$. In this circumstance, steps (4.8), (4.9), (4.10), and (4.11) incorporate an interpolation procedure designed to deal with the problem. We have not yet executed this subroutine and so cannot attest to its adequacy in handling the situation.

Statement (4.12) actually follows directly after statement (4.5), provided that the data passed the two tests. In the great majority of calculations, this will occur so, in estimating the speed of this main loop, one should not consider the delay in executing the intervening statements.

Most of the remaining instructions will be clear to the reader who is willing to follow them carefully, using Figs. 45 and 46 as a guide. They are largely derived from the application of the law of sines and the law of cosines to various oblique triangles which can be found on the figures.

A number of these statements are almost directly the same as similar statements in Mark 3 raytracing. For example, statement (4.21) is a direct equivalent and statement (4.24) is the more accurate reduction of the law of cosines, discussed at some length in the section on

Mark 3 raytracing. Statement (4.25) is the application of the parabola theorem exactly as it was used in Mark 3 raytracing.

The main loop is essentially finished with statement (4.29); the remaining statements set up the parameters for a re-entry of the main loop to calculate the next ray segment. It is necessary to calculate all the quantities listed in step (4.1) before the main loop can be re-entered. A few essential but obvious steps have been left out of this description. For example, the generation of group and phase time delays is straightforward, once we have calculated S_p and the needed refractive index average.

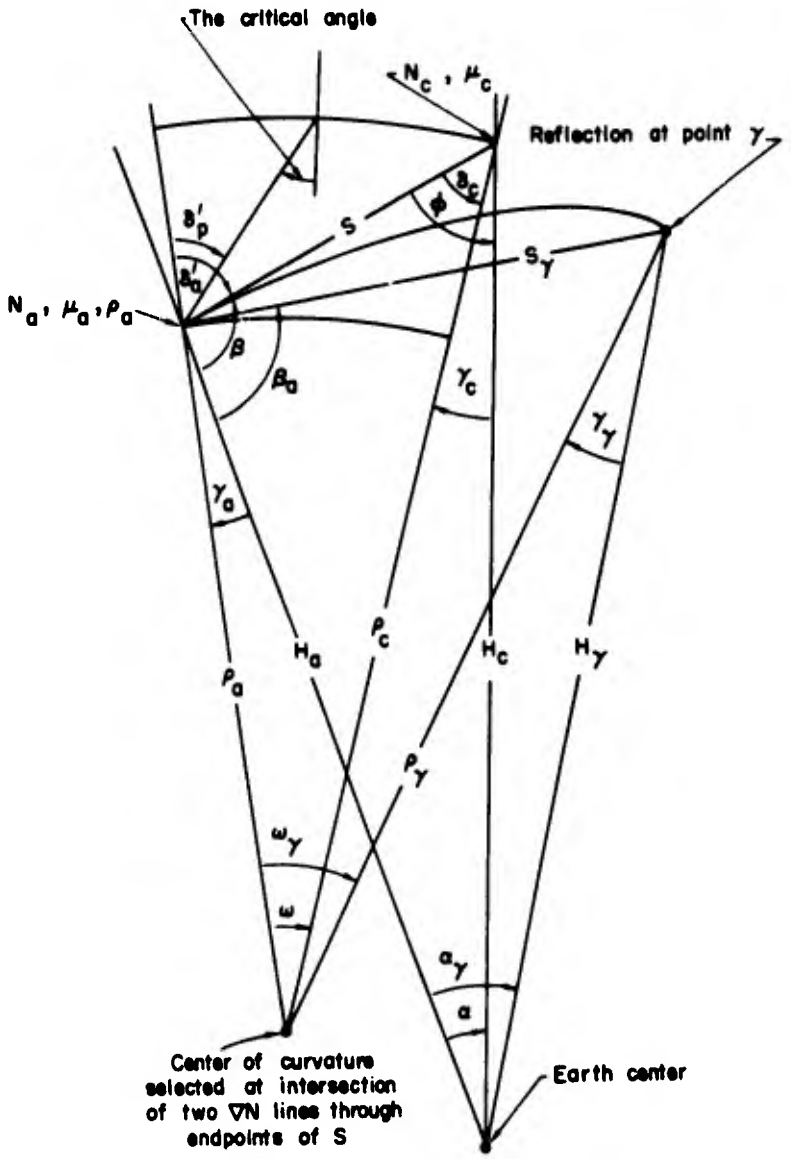


Fig. 47. LOCATING A PARABOLIC ARC TO THE REFLECTION POINT.

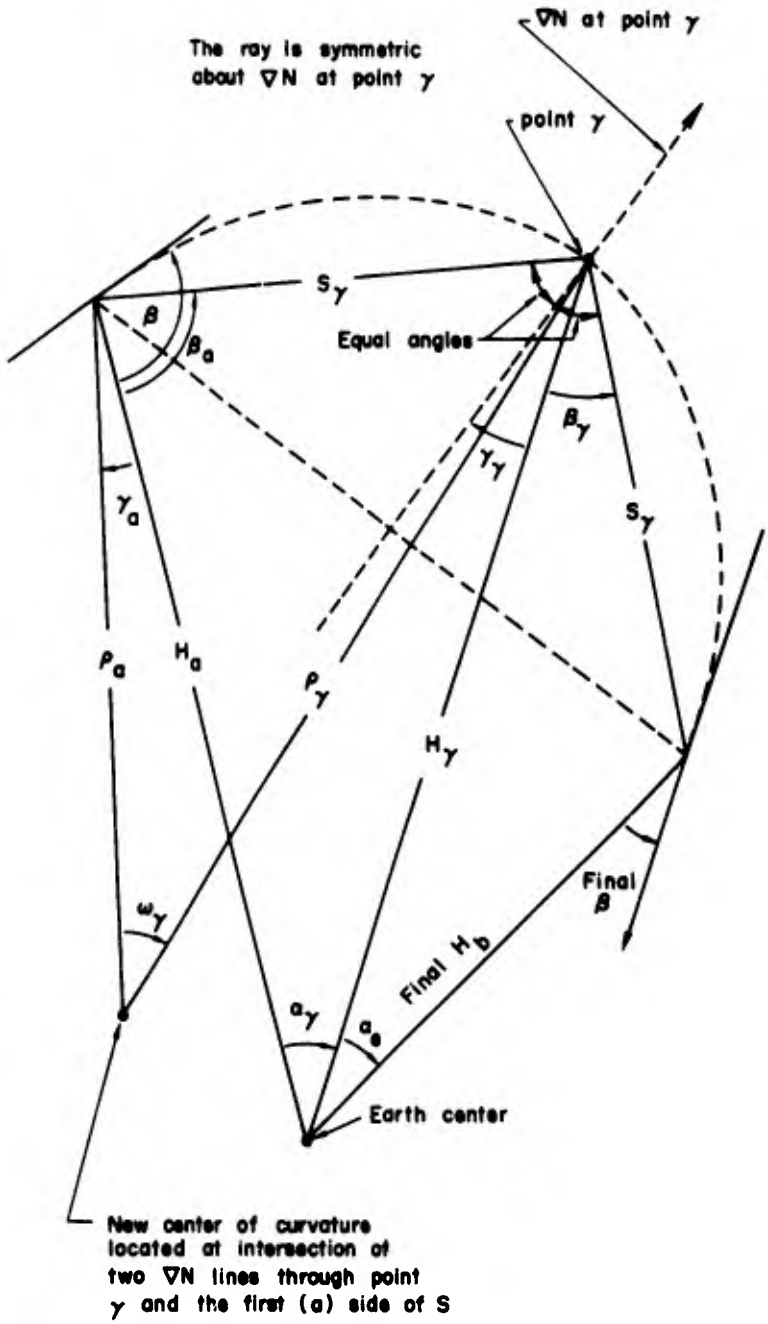


Fig. 48. PARABOLIC ARC AS THE RAY CURVES AWAY FROM ITS REFLECTION POINT.

2. The Reflection Routine

In the main loop, it is necessary to check at each step to see that $\sin \phi'_c \leq 1$, otherwise ϕ'_c is undefined because the ray reflects

at (H_c, α). Suppose $(\mu_b/\mu_c) \sin \delta_c$ $(= \sin \Delta_c')$ > 1. We would have found the following parameters already in the main loop:

$$H_c, \alpha, \varphi, \gamma_c, N_c, \mu_c, \delta_c, \omega, \rho_b, \rho_c, y, z, \dot{\rho}_b \quad \text{(plus } S, H_b, \beta, \gamma_b, N_b, \mu_b, \delta_b')$$

We must find a parabolic segment over the top and generate a sequence of steps which result in new values for these parameters, before we can re-enter the main loop. The following steps will accomplish this (refer to Figs. 47 and 48).

$$\delta_p' = \sin^{-1}\left(\frac{\rho_c \mu_c}{\rho_b \mu_b}\right)$$

$$W = \frac{\pi/2 - \delta_b'}{\pi/2 - \delta_p'}$$

$$\rho_\gamma = \rho_b + W^2(\rho_c - \rho_b)$$

$$\omega_\gamma = \frac{4(\rho_\gamma - \rho_b)}{\dot{\rho}_b(\rho_\gamma + \rho_b)}$$

$$S_\gamma = \sqrt{\left(\rho_\gamma - \rho_b\right)^2 + \rho_\gamma \rho_b \omega_\gamma^2 \left(1 - \omega_\gamma^2/12\right)}$$

$$\mu_\gamma = \frac{\left(\mu_c - \mu_b\right)\left(\rho_\gamma - \rho_b\right)}{\rho_c - \rho_b} + \mu_b$$

$$\Delta T_g = \frac{S_\gamma}{\bar{\mu} c} \;\&\; \Delta T_\varphi = \frac{\bar{\mu} S_\gamma}{c} \quad \text{where}$$

$$\bar{\mu} = \frac{\mu_b + \mu_\gamma}{2}$$

$$\beta_b = \sin^{-1}\left(\frac{\rho_\gamma}{S_\gamma} \sin \omega_\gamma\right) - \gamma_b$$

$$H_\gamma = \sqrt{H_b^2 + S_\gamma^2 - 2H_b S_\gamma \cos \beta_b}$$

$$\alpha_\gamma = \sin^{-1}\left(\frac{S_\gamma}{H_\gamma} \sin \beta_b\right) = \Delta\theta$$

Read γ_γ at $(H_\gamma, \alpha_\gamma)$

$$\beta_\gamma = \pi - \alpha_\gamma - \beta_b - 2\gamma_\gamma$$

$$H_b = \sqrt{H_\gamma^2 + S_\gamma^2 - 2H_\gamma S_\gamma \cos \beta_b}$$

$$\alpha_e = \sin^{-1}\left(\frac{S_\gamma}{H_e} \sin \beta_\gamma\right) = \Delta\theta$$

$$\beta = \pi + \alpha_e - \alpha_\gamma - \beta - 2\gamma_\gamma$$

Read N_b, γ_b at (H_b, α_e)

Test $N_b < f^2/80.6$, find μ_b

$$\bar{\mu} = \frac{1}{2}(\mu_b + \mu_\gamma)$$

$$\Delta T_g = \frac{S_\gamma}{\bar{\mu}c} \;\&\; \Delta T_\varphi = \frac{\bar{\mu} S_\gamma}{c}$$

Re-enter Main Loop

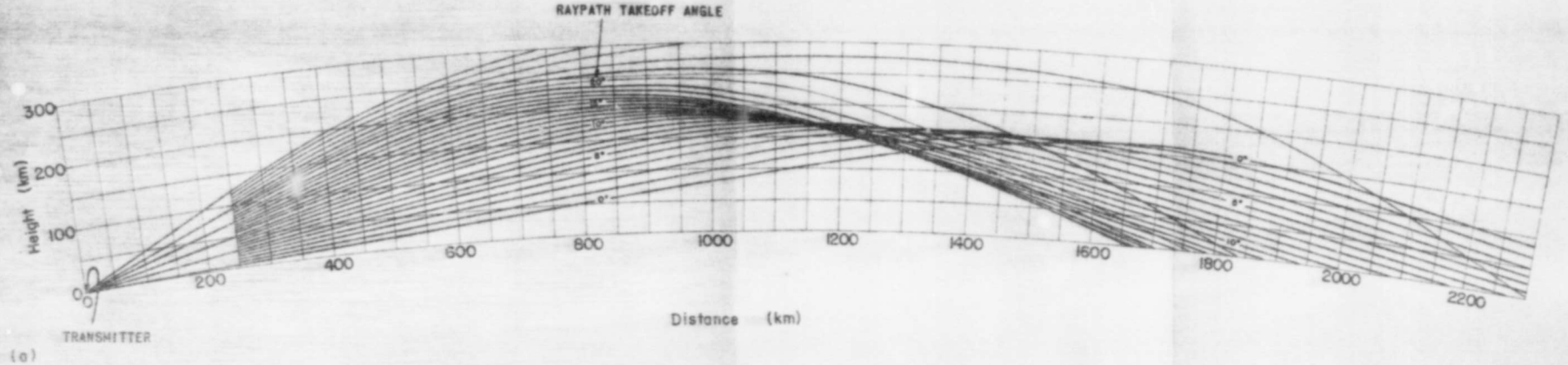
A Simple Proof of the Parabola Theorem. After this report was printed, Mr. Taylor Washburn examined a copy and pointed out the following to the author:

A parabola may be described without loss of generality by $h = ar^2$.

Then, $\dot{h}(r_1) = 2ar_1$ and $\dot{h}(r_2) = 2ar_2$, and their average is $a(r_1 + r_2)$.

The chord slope is $\frac{\Delta h}{\Delta r} = \frac{ar_2^2 - ar_1^2}{r_2 - r_1} = a(r_1 + r_2)$.

This shows that the chord slope equals the average of the two parabola slopes measured at the chord endpoints. On pages 76-77 this is shown for the same parabola translated in the coordinate system, represented by the formula $h = ar^2 + br + c$, but obviously this property of parabolas is invariant under coordinate translation so that the more simple proof given above is superior. (It is included here, out of place in the report because this page was originally blank and we did not want to reprint the existing pages.)



(a)

(b)

Fig. 49. COMPARISON BETWEEN A HAND-PLOT AND A MACHINE PLOT. (a) Hand drawn and (b) machine drawn.

BLANK PAGE

VI. PLOTS OF RAYS CALCULATED BY VARIOUS PROGRAMS

In the author's previous report on raytracing [31], the most widely appreciated features were the plots of the rays. Thus, it seems appropriate to include here a number of other examples of plotted rays. These will be used to illustrate various technical points associated with ray propagation and its simulation through raytracing.

A. Hand Plots Compared to Machine Plots

To emphasize the full value of machine plotters, Fig. 49 shows a comparison between our best hand-drawn example and a good machine-drawn ray plot. Figure 49a was drawn in 1961 shortly after the author devised his first program. As "input data" the draftsman was given a tabular list of successive locations of each ray. This took a long time to draw, because it is inherently difficult to draft lines of uniform weight and spacing. If the author's memory is correct, this drawing took two man-days and involved much rework to remove nonuniform features. Only two such drawings were attempted before we became convinced that manual plots were impractical.

At Stanford, plotting is done with Calcomp equipment and with an ink pen. Ballpoint pens are more reliable but the plots are markedly inferior. Two different plotters provide different levels of fine-scale detail. Our best quality is obtained when the fine-detail machine draws a large plot, but this is also the most expensive mode of operation so it is seldom used. Figure 49b shows one such plot which was made for a special purpose that justified the plot cost of approximately $20. It should be noted that even this cost is far less than the drafting cost of part (a). (For the convenience of those who have access to the author's previous reports [16,32] Fig. 49b is computed in IID 245 at 24 MHz with $\Delta\beta = 0.5°$.)

B. High Frequency Rays Computed with the Magnetic Field

It is easily possible to change the color of plotted rays by simply instructing the plotter operator to change pens to the appropriate color

at the appropriate time. We have taken advantage of this facility by incorporating a convention whereby all ordinary rays are drawn red, all extraordinary rays are drawn green, and all rays computed without the field are drawn in black. This is convenient and we recommend it to others who do this work. However, it has a disadvantage in a publication because it is very expensive to photograph and reproduce a color picture using the techniques available in the technical reporting facilities. It is necessary to give the original three-color plots to draftsmen who trace each color into a separate black drawing. These tracings are then used in a photographic process to produce the color figures shown here. In most applications, these plots would probably be rendered in black for publication, but here it is intended to convey the utility of the 3-color method which is inexpensive in its original form.

1. The Azimuth-Dependence of the Geomagnetic Effect

In mid-1965, using the Mark 8 program, we investigated the nature of the ray-splitting effect of the magnetic field. The main consequence of this study has been a realization that the splitting can usually be neglected or else it can be accounted for by means of approximations. Now, four years later, the cost of including the magnetic field is reduced and so it may be practical to include the field in many calculations where it was previously unwarranted.

Figure 50 shows a globe on which four paths are indicated. In order to check the magnetic azimuth dependence of the splitting, we computed rays at 4 MHz along each of these paths. The result is shown of Fig. 51, where it is seen that north-south rays are split about twice as much as are the east-west rays. The rays on a 45° azimuth exhibit an intermediate effect, as one might expect.

2. The Frequency-Dependence of the Effect

Since the azimuth study contained no surprises, the frequency dependence tests were conducted at a single azimuth. We chose the Seattle-Laredo path at about a 45° azimuth, the same as was used to generate the fourth plot on Fig. 51. On this azimuth, rays were calculated

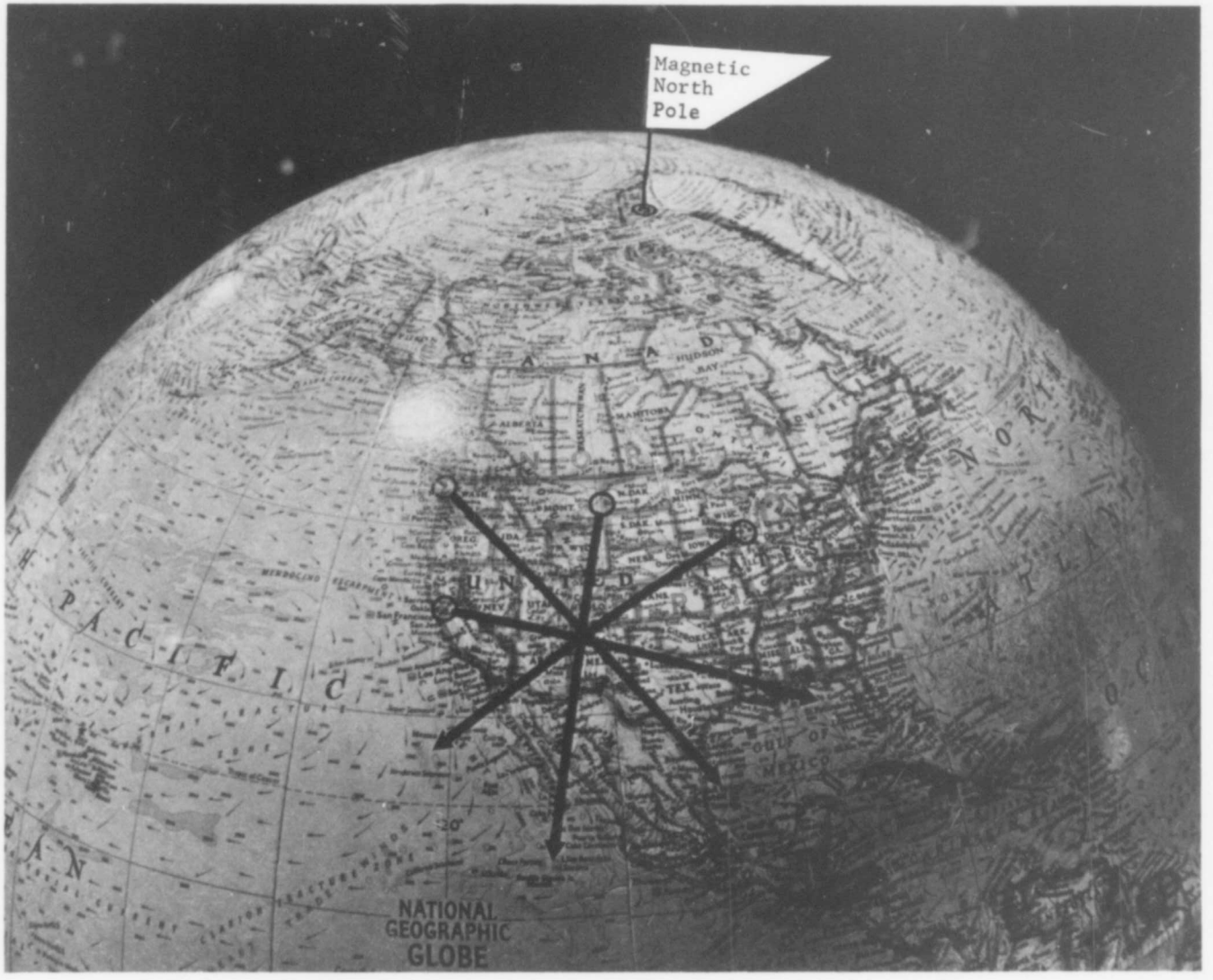


Fig. 50. FOUR SIMULATED HF PATHS RELATIVE TO THE MAGNETIC POLE.

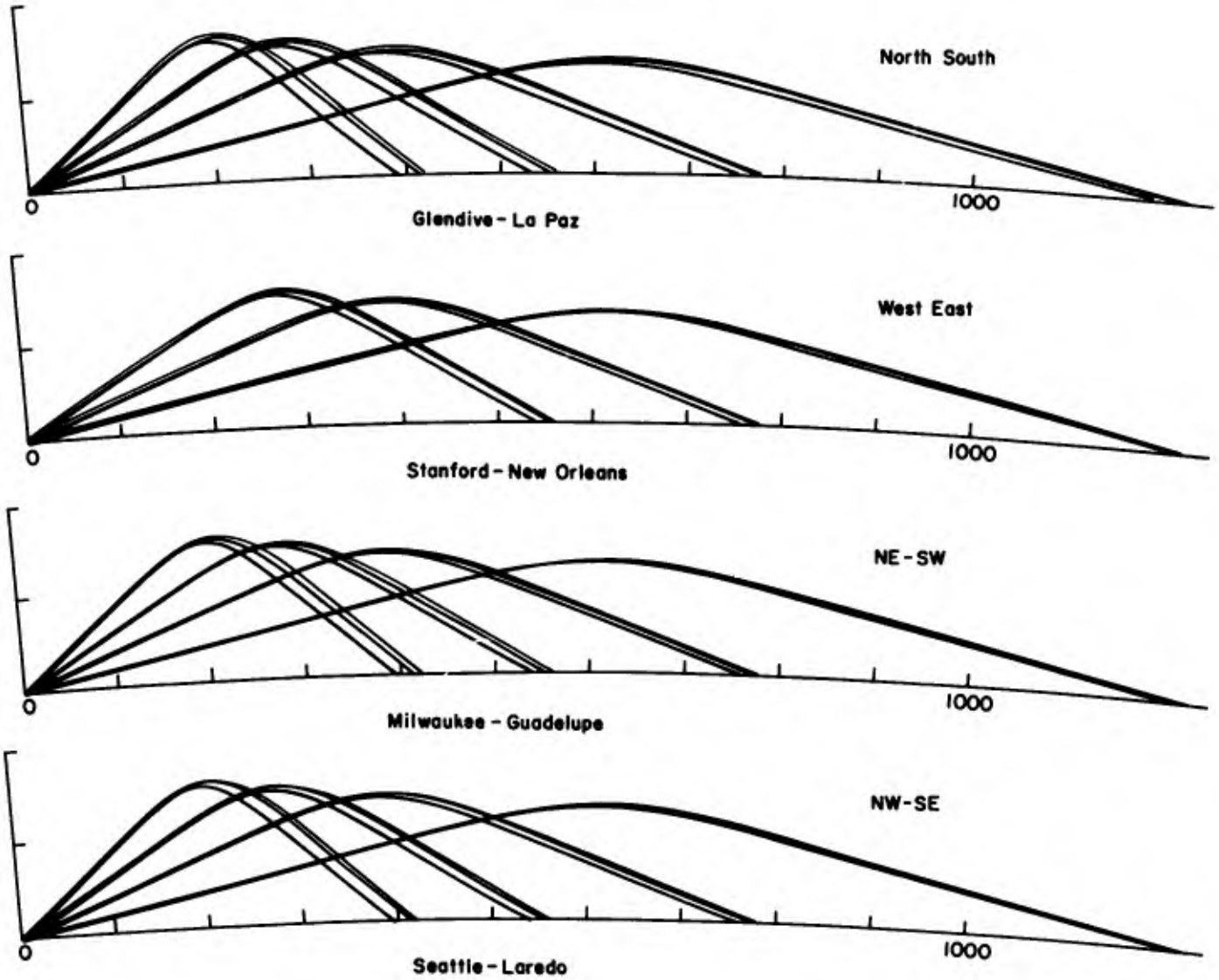


Fig. 51. MAGNETIC SPLITTING AS A FUNCTION OF AZIMUTH.

at three or four takeoff angles per frequency from 1 MHz up to the maximum possible sky wave frequency, which was 30 MHz in this model ionosphere (IID 165). The resulting rays are shown in Fig. 52, but in the interest of economy the long-range rays above 18 MHz are not included.

It is perhaps most interesting to notice that the separation of the rays at their termination, measured in meters normal to the rays, does not have much frequency dependence above 4 MHz. As the frequency increases, the rays diverge less rapidly but they also travel farther, so there is some compensation. There is an effect, of course, but it is less than the author expected to find. It should be added that these answers, computed with the Mark 8 program, are not as accurate as could now be obtained, using the new ESSA program [25,33].

3. The Cost of Such Calculations

It is difficult to provide good cost data because so much depends on the chosen mode of operation. In the example of Fig. 53, the step length was 1 km but it could have been set to 10 km and the costs would have been much less. Unfortunately, there is no way of knowing in advance just how long to set the step, and so one must rely on experience with each particular program. The accuracy deteriorates with increasing step length, and so one is forced to decide just how much accuracy should be purchased.

As is shown in Fig. 53, the cost per ray was $10. The newer ESSA program is said to be superior and less expensive. Also, it should be pointed out that $18 of the cost of Fig. 53 was incurred compiling the program, a step which is only needed when modifications have been made. The rays actually cost only $4 each and, if a plot had not been made, the cost would have been nearer $2 each. For comparison, it should be pointed out that similar rays computed by Mark 1 cost about 10¢ and those computed by Mark 3 cost about 0.3¢. All of these comparisons reflect the author's judgment, because it is difficult to isolate a repeatable measure of cost.

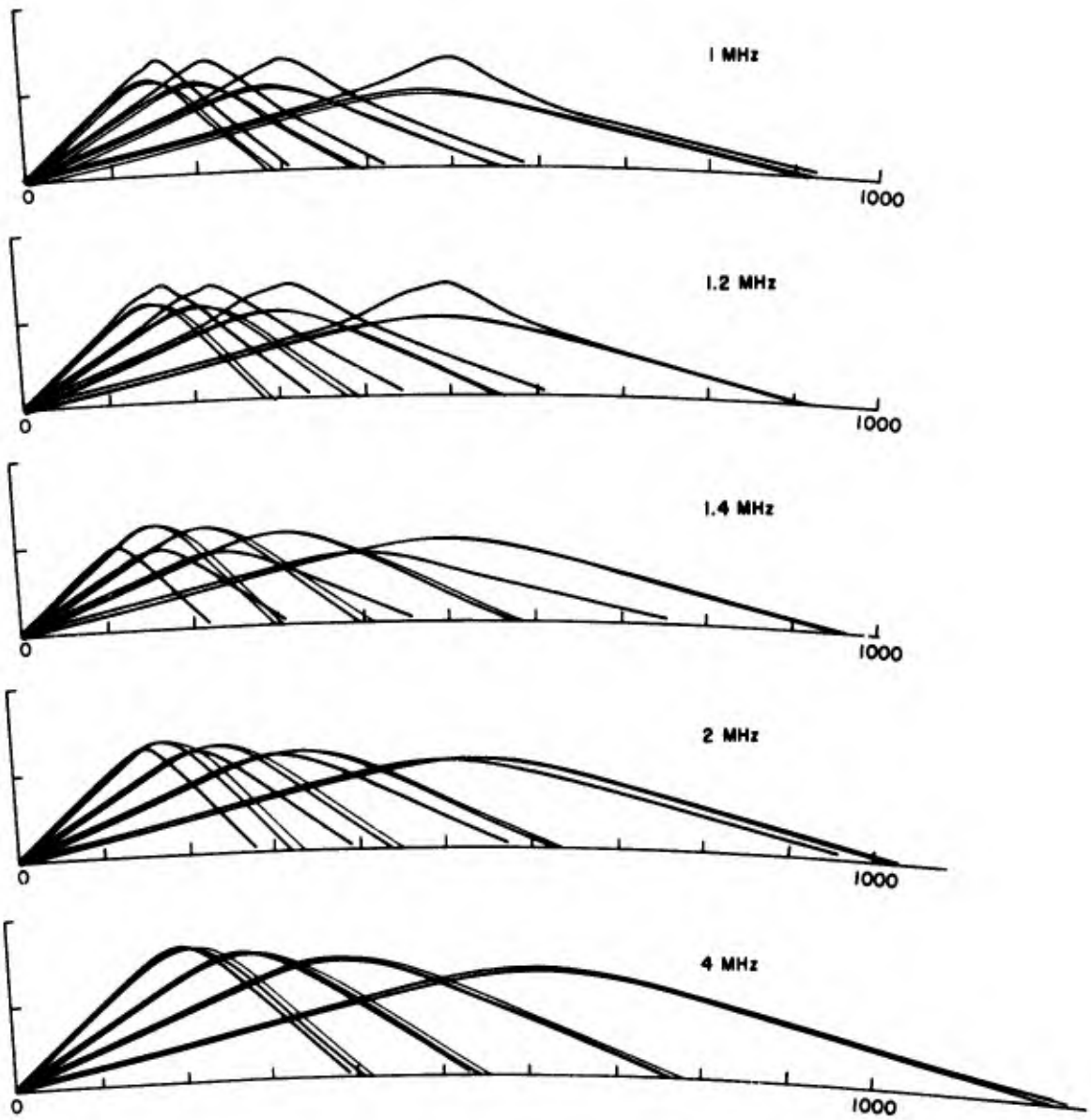


Fig. 52. MAGNETIC SPLITTING AS A FUNCTION OF FREQUENCY IN IID 165.

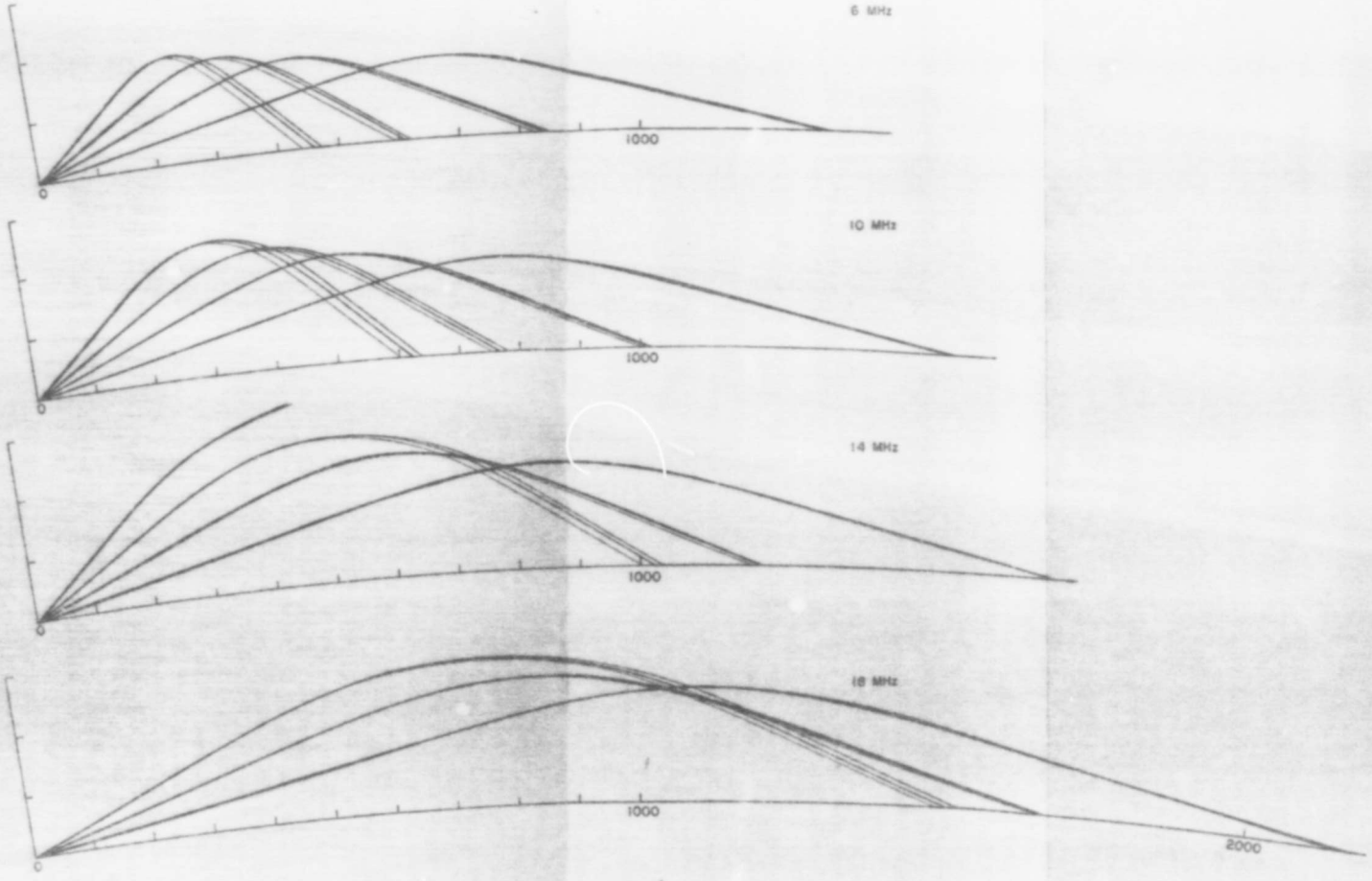


Fig. 52. CONTINUED

	IID	FREQ KC	HOP	BETA DEG	PSI DEG	GROUP TIME MS	PHASE TIME MS	RANGE KM	HT KM	INUMB
1	165	1500.	1	10.00000	10.29754	3.4532616	3.4208205	1003.3810	95.85496	0
2	165	1500.	1	10.00000	10.28303	3.4666041	3.4341028	1007.3449	95.84192	0
3	165	1500.	1	10.00000	10.32979	3.1330400	3.0552109	896.8696	85.05962	0

Time & Cost of Computation

Item	Min	Sec	Dollars
7090 Compile	4	20	18.00
7090 Read In			
7090 Compute	2	24	10.00
Plotter Work	4		2.00
Total Cost of this job			$30.00

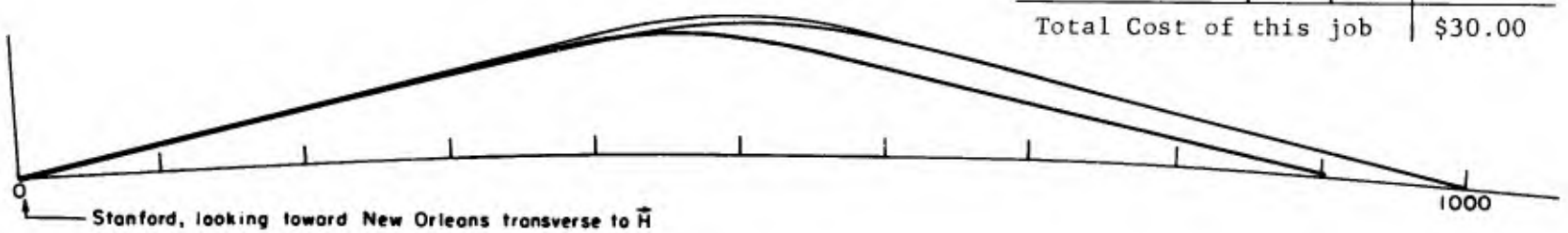


Fig. 53. EXAMPLE OF RAYSETS, RAYS, AND COST DATA.

C. A Focusing Phenomenon Discovered by Means of Computed Rays

Sometimes ray patterns reveal patterns of propagation which exist in the real world but which are not apparent in real data. An excellent example of this application of rays is provided by the discovery that the ionospheric structure between layers causes a focusing which is analogous to skip focusing caused by the layers themselves [34]. The author first noticed this effect in synthetic oblique ionograms and then in real ionograms. The effect was then traced back to computed rays, such as those illustrated on Fig. 54. In part (a) of the figure, rays are shown in a two-layer ionosphere (IID 213); these are skip-focused at 1660 km by the E layer and at 1540 km by the F layer. When electrons were added between layers (IID 215), the rays changed to reveal the pattern shown in

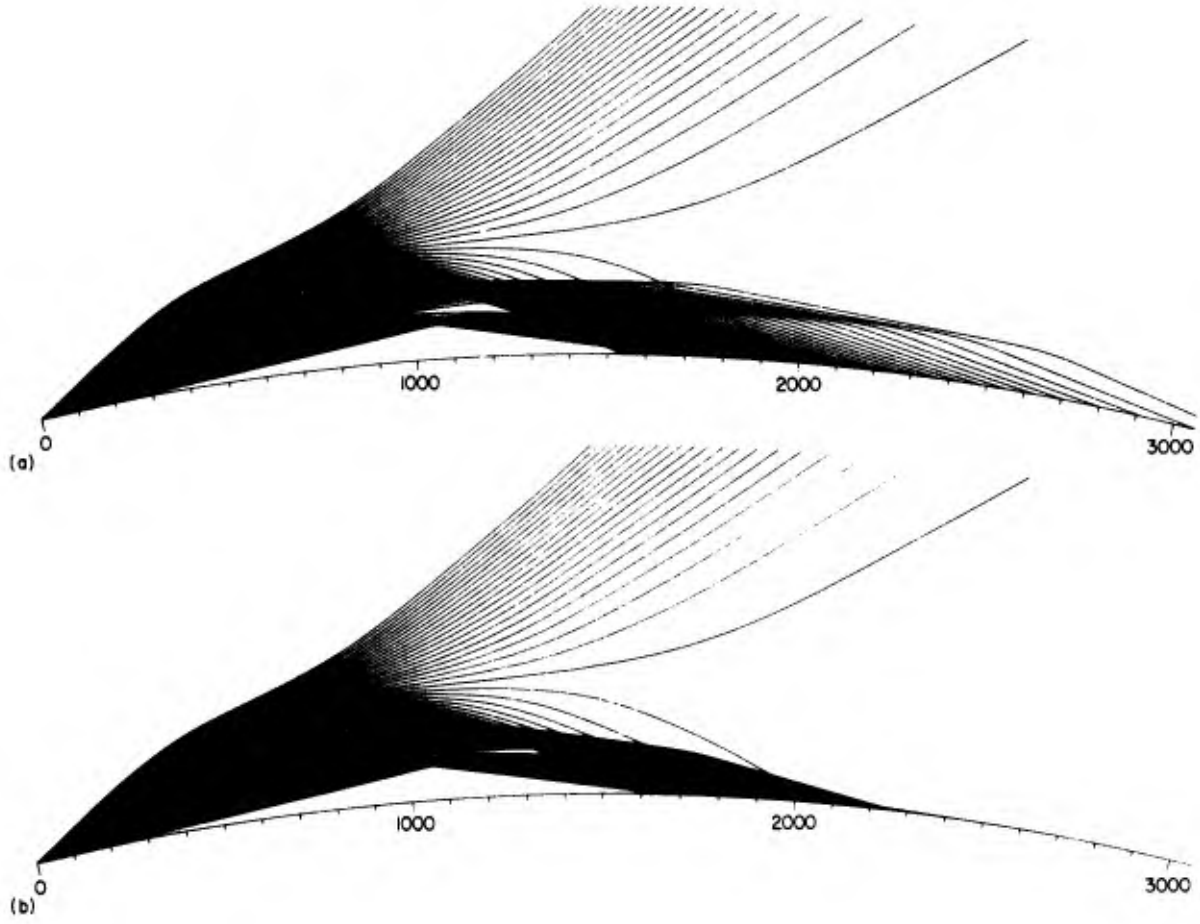


Fig. 54. RAYS WITH AND WITHOUT FOCUSING CAUSED BY THE ELECTRON DISTRIBUTION BETWEEN LAYERS. (a) Rays without the focusing and (b) rays with interlayer focusing.

part (b). Here, a heavy focusing of rays at 2250 km is shown. This is caused by the interlayer electrons. The effect and these same rays were discussed at length in the reference cited. After finding the effect in rays, it was further tracked into sweep-frequency backscatter where the phenomenon was found to explain a feature of the data which had previously been a puzzle.

The frequency dependence of this focusing may be clearly seen in Fig. 3 (of this report) which showed rays calculated throughout the HF band in ionosphere IID 215. At the lower frequencies, the rays are constrained below the E layer. As the frequency increases, some of the higher-angle rays spray upward toward the F layer, but they are confined within a channel along the sides of which the energy is concentrated. The lower side of this channel strikes the ground and is called skip-distance focusing. The upper side constitutes the interlayer focusing. As the frequency continues still higher, the effect disappears because all rays penetrate through the interlayer electrons and are controlled primarily by the F layer alone.

This frequency dependence of interlayer focusing can be shown in a more compact but less intuitively obvious manner by means of a reflectrix family. This display was used in the reference cited because of its brevity, and the rays were not included because it is difficult to present ray details on the small pages typical of journals. This illustrates one of the disadvantages of ray plots drawn to scale; that is, they are very long but not very high. Some authors choose to distort the rays in order to produce a plot of more practical shape, but this always distorts the angular relationships and thus destroys some of the value of the plots.

D. Variations in the Plot Coordinates

A raytracing program which can follow rays in the circularly disposed ionosphere can be easily modified to follow rays in other media of similar character. Often, all that is needed is a different plot format. Figure 55 illustrates one such modification wherein the rays are emanating from a point which appears to be 50,000 km from the earth's center.

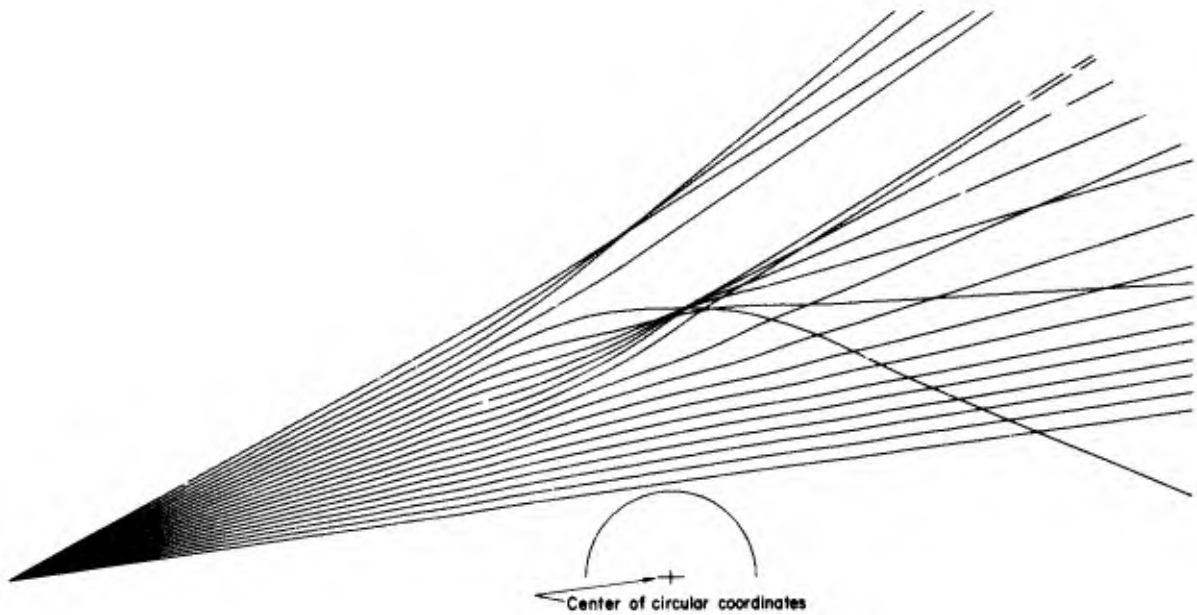


Fig. 55. SLIGHT MODIFICATION OF THE RAY PLOTTING COORDINATES.

In fact, the objective of this display is to aid in the study of phenomena which may be much smaller or much larger than 50,000 km. The "earth" could represent a star, for example, and this plot might show the track of radio waves in its corona. Alternatively, the plot center might represent a small idealized "blob" in the ionosphere of only a few meters dimension.

For some types of rays, it is desirable to have the source at an infinite distance, that is, to have the rays enter parallel. This modification can be programmed in a straightforward manner and the results appear like the example shown in Fig. 56. The change from conventional programs to this format is time-consuming but not difficult. In the changeover, it is desirable to redefine rayset parameters to fit this new viewpoint, but this is easily done.

One very useful program modification is the incorporation of the ability to originate rays above the earth's surface. Figure 57 shows examples of this, one with and one without time delay ticks at 0.5 msec

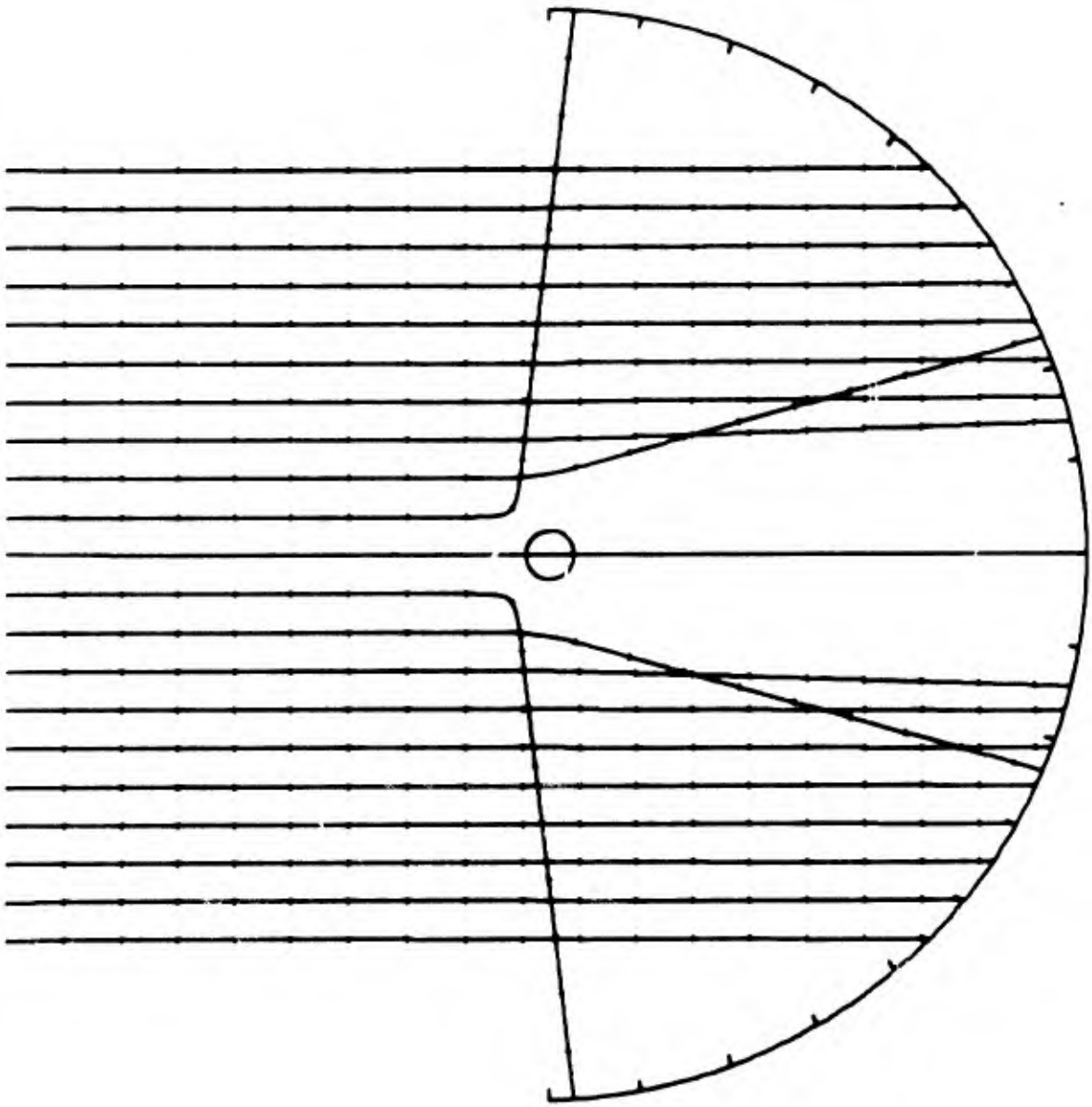


Fig. 56. FURTHER MODIFICATION WITH RAYS ENTERING PARALLEL.

intervals of group delay. The story behind these plots illustrates an important point: We provided a raytracing program to a scientist on the east coast. A few weeks later, we received a call saying that the plotted rays were uselessly erratic. On our request, the entire input-output set was sent to us and we found that the ionsopheric model being used consisted of a Chapman layer with electron densities evaluated to only four significant digits. To an engineer, trained to approximate,

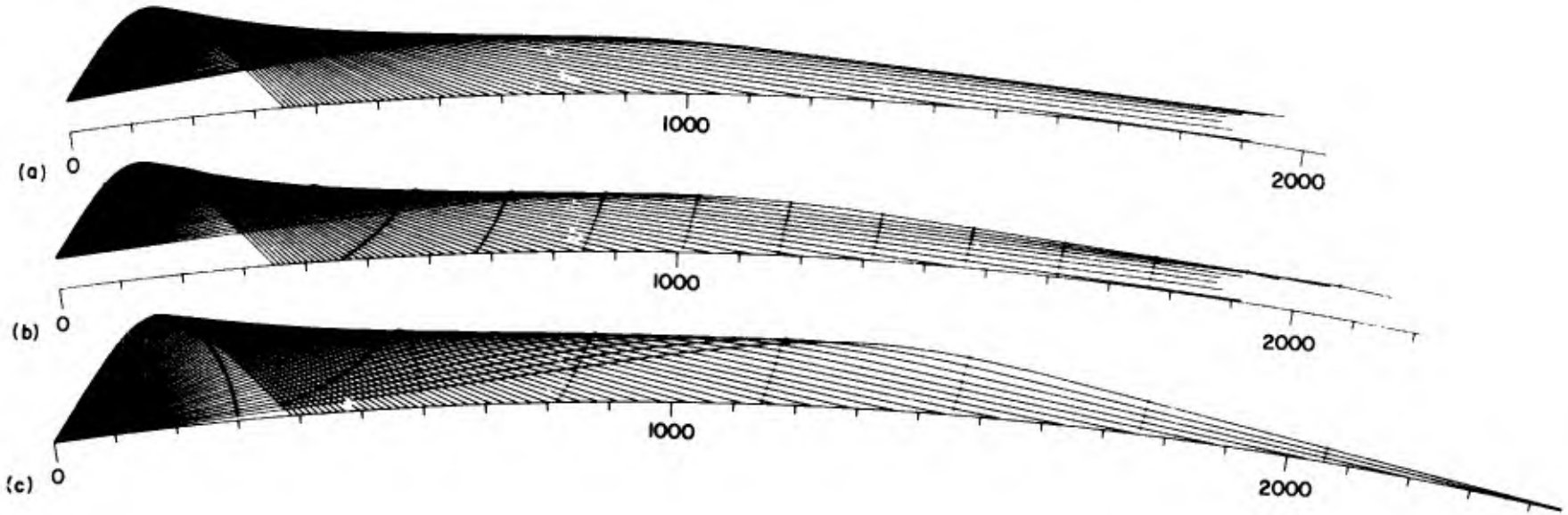


Fig. 57. VARIATION OF RAY ORIGINATION ALTITUDE. (a) Rays launched at 50 km, (b) same rays with group delay ticks each at 0.5 msec, and (c) same rays launched from the surface with group delay ticks each at 1 msec.

this seems reasonable. However, the amount of error introduced by round-off at this level caused the ray families to appear ragged and also made it virtually impossible to calculate power density through the use of ray spacings. The moral is: use all the significant digits you have when you set up input data for a computer. Unlike the situation in manual processing, it is almost free to have more digits in a computer process and it avoids the introduction of errors which must later be tracked down.

The rays in parts (a) and (b) of Fig. 57 terminate in midair because the computer was instructed to stop at the first perigee. It has been found worthwhile to count perigees and also to separately count ground encounters. The program user can exercise separate control over the maximum desired number of perigees and ground encounters.

For comparison, Fig. 57c shows the same calculation with one exception, the transmitter is on the ground. Here is clearly shown the time history of the signal distribution in space as the energy starts upward, only to turn around and return to earth as a simple straight front.

E. Rays of Unique Appearance

Often there exists a propagation mode in which the calculated energy density is so low that for all practical purposes, the mode does not exist. However, in digital ray calculation, signal strength is analogous to the probability of finding rays. In other words, if a mode is weak, you probably won't find a ray in that mode, but you might! This sometimes produces a few surprises. Figure 58a shows an example: a ray that travels alone for 5900 km in one hop. This mode is very weak; a change in takeoff angle of 1/100 degree will decrease the range of the ray so that it appears to be part of the E layer family which lands at 1800 km or part of the F layer family that lands at 3300 km. (This was calculated at 5.8° in IID 245 at 23 MHz.)

On Fig. 7 of Ref. 31, a trapped mode was shown between the E and F layers. The 2 MHz rays are recalculated in Fig. 58b, and the trapped 6 MHz rays are given in Fig. 58c. These modes should actually exist and, in fact, they should contain reasonable radio energy density. On these calculations, notice how the Mark 1 programs properly indicate the

shortening length of the oscillations. These rays were launched from the nightside of the earth, crossing the terminator as they entered the ionosphere. The E layer thus increases in density with increasing distance. These rays are able to penetrate the E layer going up, but thereafter they cannot penetrate again, and they become trapped. Subsequently, the valley between layers fills with electrons until the 2 MHz rays can no longer propagate; thus, they end with a near-vertical oscillation, caused by convergence of the isodensity lines. (This is much like the turning back of trapped electrons in the magnetosphere caused by the convergence of geomagnetic field lines as the charged particles approach the earth.)

F. Phase and Group Fronts without the Rays

When the ionospheric model is realistically complicated and the fronts are calculated, it is often difficult or impossible to discern the front locations by inspecting the final plot. At first, the author attempted to overcome this problem by making separate calculations of subsets of each ray family. To illustrate this process, Fig. 59 shows the rays used to produce the two-layer phase fronts of Fig. 6 which were computed in IID 166. Figure 59a shows an entire ray family consisting of all rays from 0 to 50° in 2° increments. It would be very difficult to draw the fronts without further clues. To extract these clues, the three other plots in Fig. 59 were generated. Part (b) shows rays from 0 to 8° in 0.4° increments; part (c) shows the next higher rays from 8 to 10° in 0.1° increments; finally, part (d) skips the obvious parts and shows just the Pederson rays from 44 to 46° in 0.1° increments. With these aids, it is clear where the fronts lie.

The preceding plot technique works well but it is laborious. A better way is illustrated in Fig. 60. The computer is instructed to remember the location of each tick. Then after the ray drawing is complete, the computer draws only the ticks. The removal of the rays leads to such clarification that it is usually easy to follow the fronts on the tick-only drawing. With this guide, one can draw the fronts on the ray plot, parts (a) and (c) of the figure. Figure 60 was computed from IID 215, a

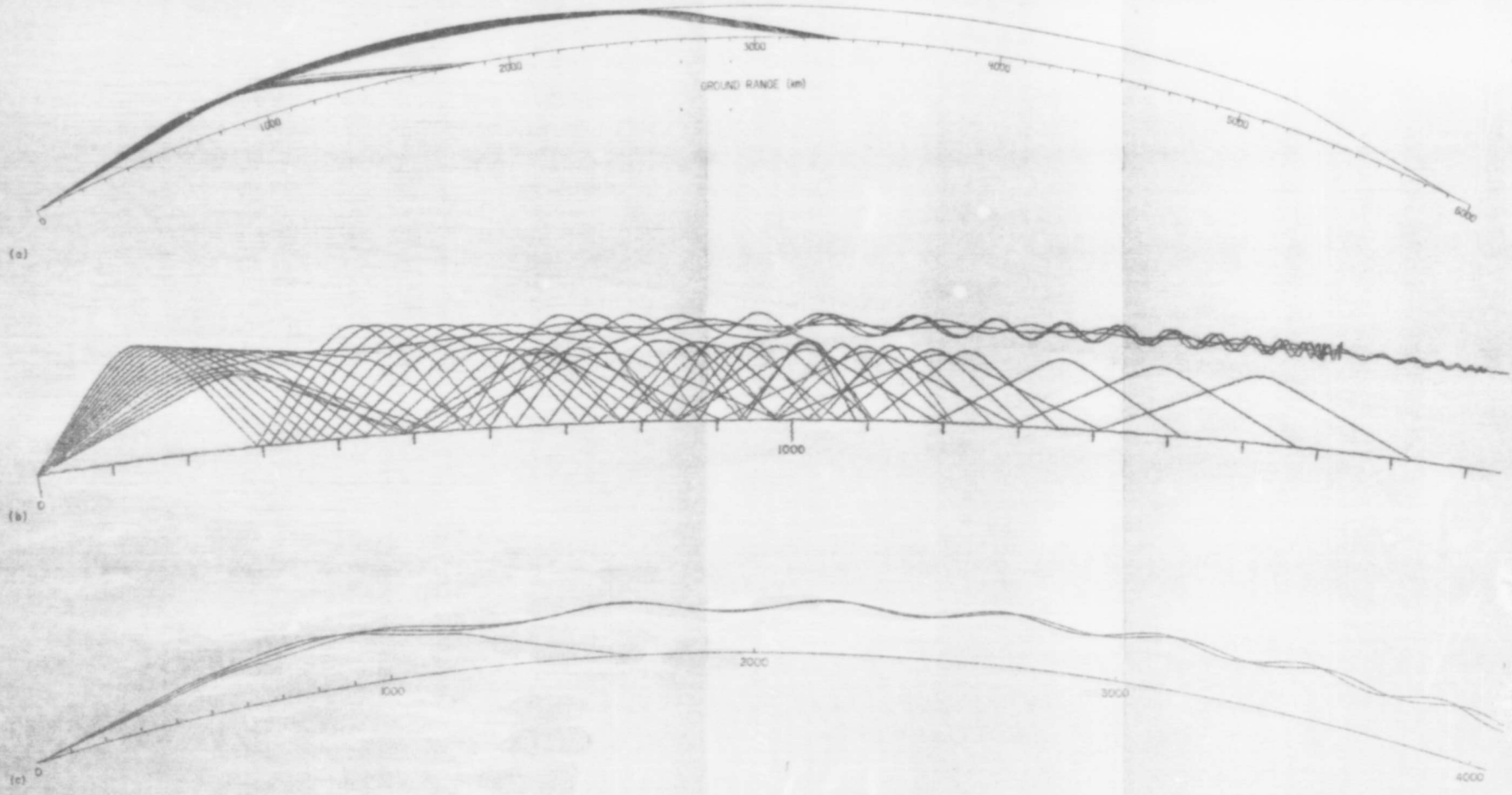


Fig. 58. ANOMALOUS RAYS WHICH TRAVEL A LONG DISTANCE. (a) Fortuitous bending which just matches earth curvature, (b) trapping between E and F layers at 2 MHz, and (c) the same trapping at 6 MHz.

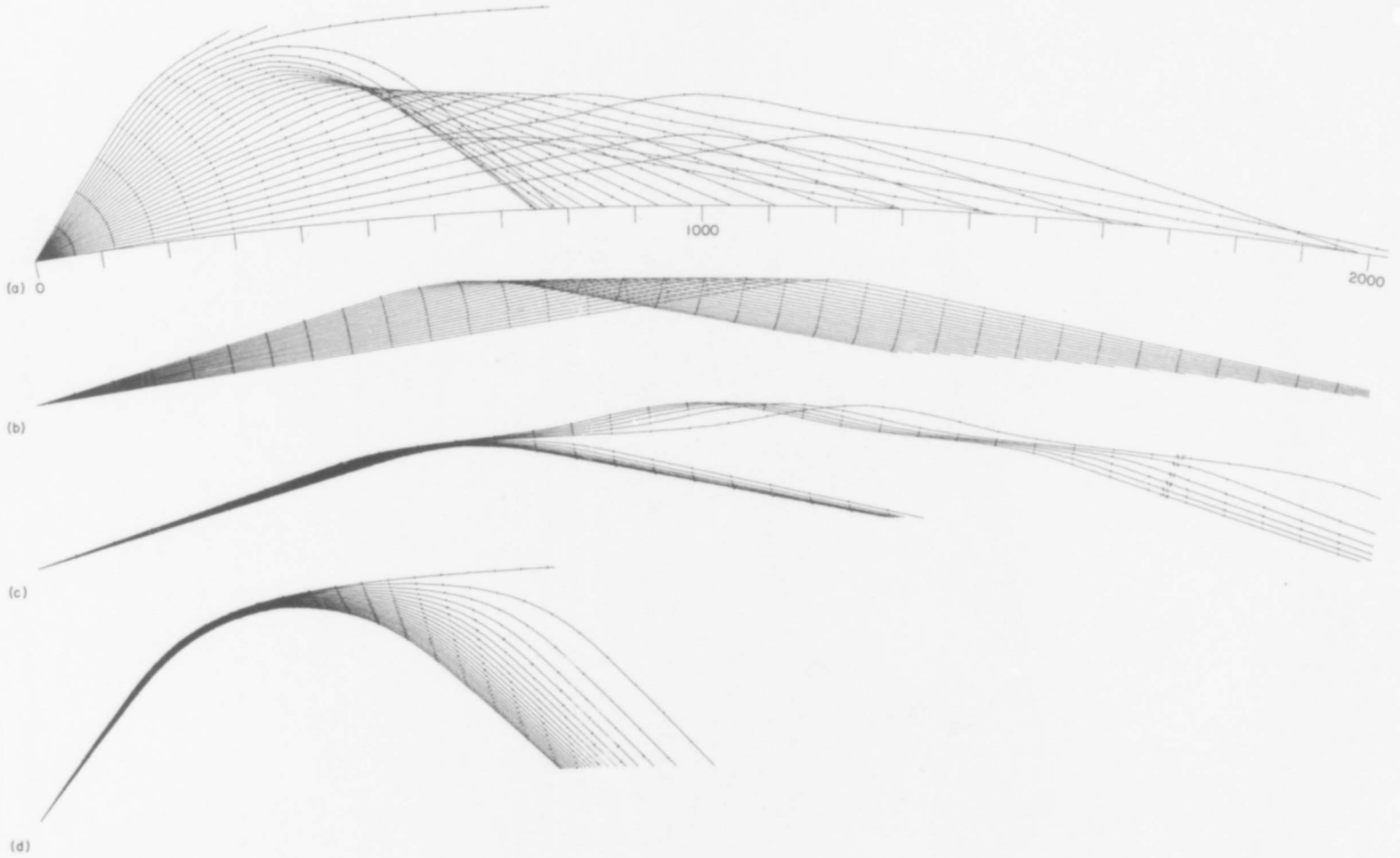


Fig. 59. FINDING THE FRONTS BY COMPUTING SUBSETS OF THE RAYS. (a) An entire set of rays, (b) subset from 0 to 8°, (c) subset from 8 to 10°, and (d) subset from 44 to 46°.

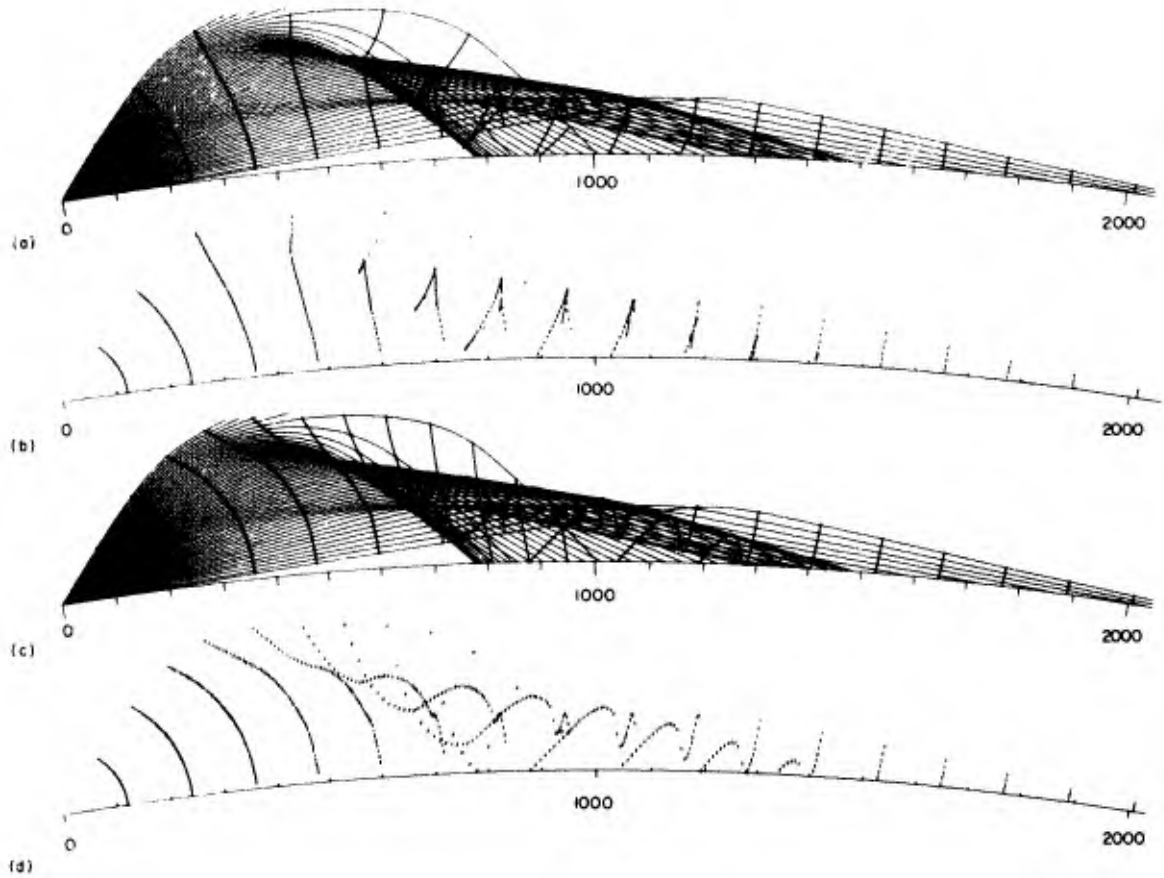


Fig. 60. FINDING THE FRONTS BY PLOTTING THE TICKS SEPARATELY, USING IID 215. (a) Phase fronts hand-drawn amid rays, (b) computer-drawn phase fronts used to produce part (a), (c) hand-drawn group fronts amid rays, and (d) group fronts.

two-layer ionosphere. Figure 61 has been made with a three-layer ionosphere, IID 217. (Both are calculated at 12 MHz.) With three layers, the fronts are considerably more complex and the need for the tick-only plot is more readily apparent. It should be noticed that one also derives valuable clues from the fact that each tick is perpendicular to the ray with which it is associated. The author did not fully appreciate the degree of this advantage until he happened to see some plots made with a small "x" at each location, rather than a perpendicular tick. The latter were far more useful when the pattern was not obvious.

The detailed structure of some fronts is of surprising complexity. Figure 62 shows an enlarged view of the signal distribution of Fig. 61c,

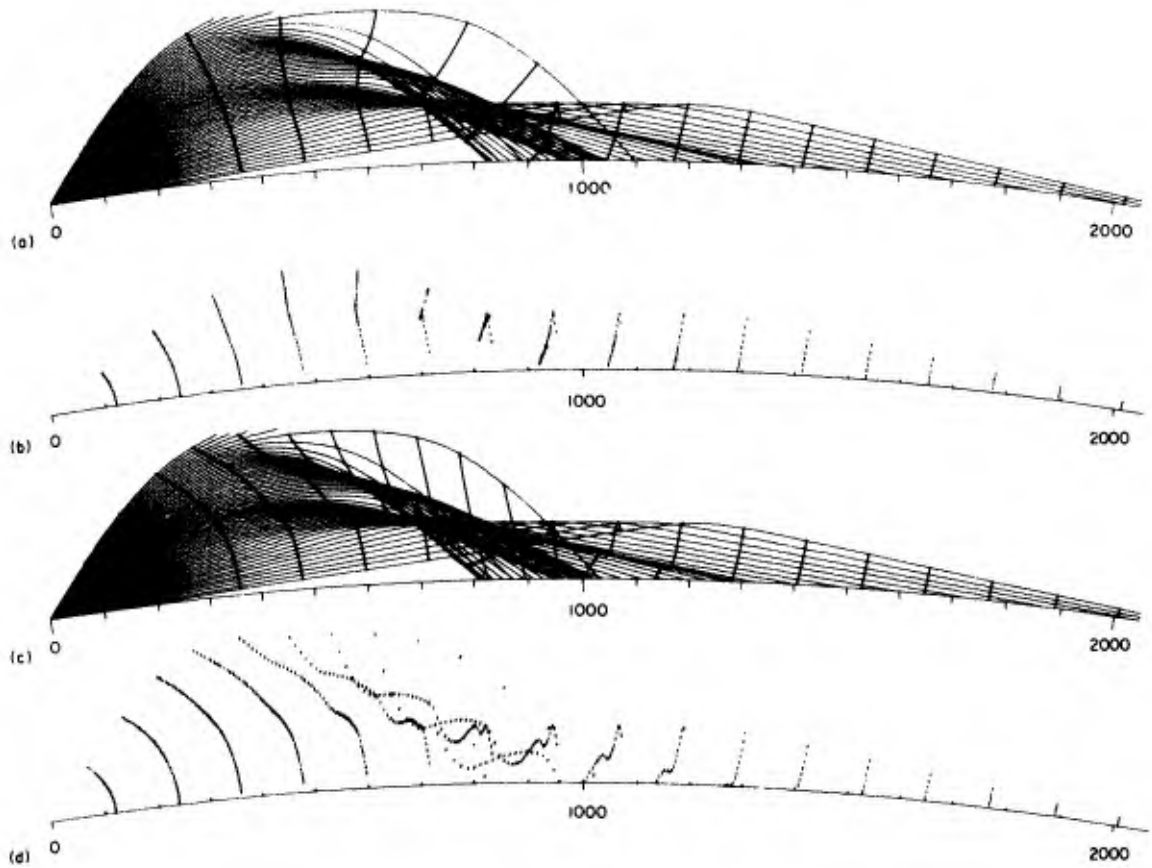


Fig. 61. ANOTHER EXAMPLE OF FRONT TRACING IN A 3-LAYER IONOSPHERE, IID 217. (a) Phase fronts hand-drawn amid rays, (b) computer-drawn phase fronts used to produce part (a), (c) hand-drawn group fronts amid rays, and (d) group fronts.

about 3 msec out from the origin. The most complete but complicated front is the 2.3 msec line, labeled at three points on the figure. It would be difficult to follow this tick pattern if it were mixed in with the rays. It would probably be impractical to calculate such a complex pattern without the aid of a computer. Yet, by using these new methods, such calculations can be carried out readily and reliably.

G. Phase and Group Fronts below the Critical Frequency

The fronts shown on Figs. 5 and 6 were calculated in IID's 165 and 166 at 12 MHz. In each model, the F layer critical frequency was 9 MHz. It is of interest to see the qualitative difference in the front

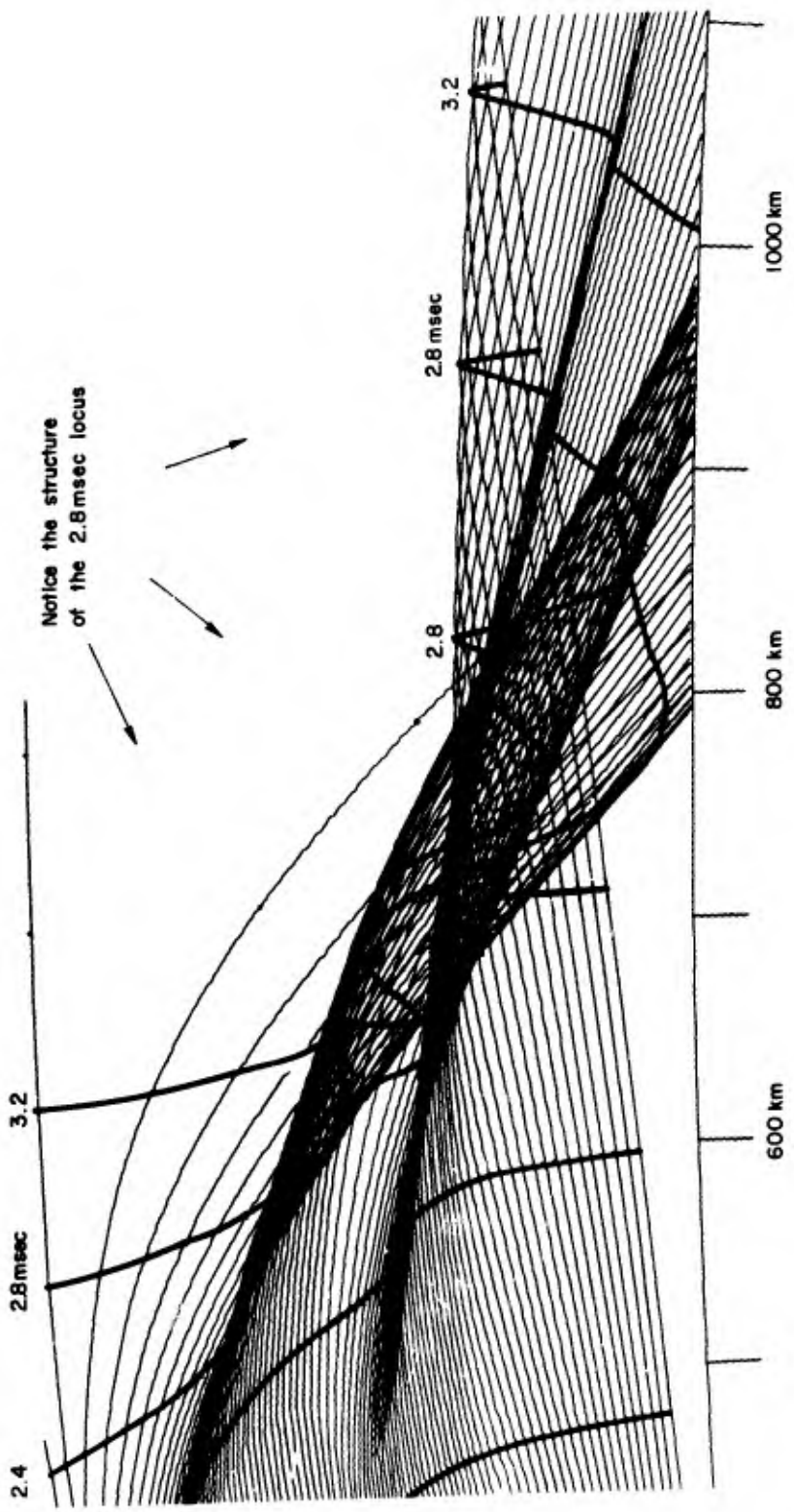


Fig. 62. ENLARGED VIEW OF THE GROUP FRONTS IN IID 217.

distributions when the rays are all confined below the F layer because of operation below the critical frequency. To illustrate this, Figs. 63 and 64 show 8 MHz counterparts of Figs. 5 and 6.

H. Ionospheric Models Used to Produce Plots

The plots presented have been computed almost exclusively with a set of twelve ionospheric models called the "nova ionospheres" by the author. For the sake of completeness, six of those used are given in the form of electron density-height plots on Fig. 65. In each case, the maximum electron density is 10^6/cc which corresponds to a critical frequency of about 9 MHz. IID's 166, 213, 215, and 217 have E layers with f_c = 3 MHz, but IID 245 has a dense E layer with f_c = 5 MHz. IID 217 has an intermediate F_1 layer with f_c = 6 MHz. The detailed structure of these was described in great detail in Ref. 16. Additionally, some plots have been computed from IID 011 which was a single Chapman layer much like IID 165. The tilted terminator ionosphere which produced E-F trapping was IID 006, described in detail in Ref. 31.

One unexpected development at this laboratory has been the continuing utility of the twelve nova ionospheres for computing additional rays in new applications. Also, the existing raysets have been kept for many years and they continue to serve diverse and unanticipated needs. The ready availability of files of existing calculated data has served many times as an acceptable substitute for ray-related data which would otherwise have to be calculated anew.

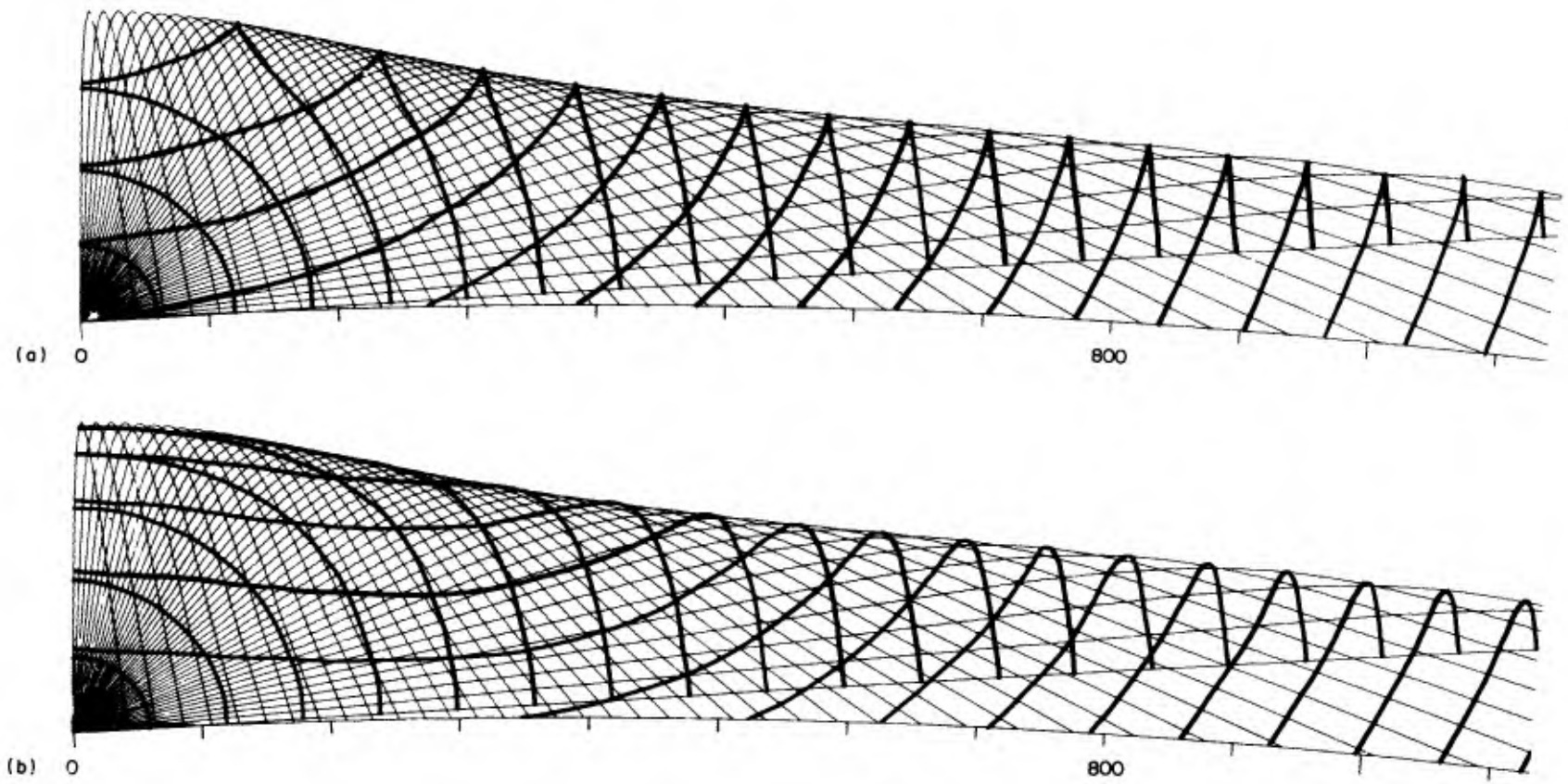


Fig. 63. 8 MHz PHASE AND GROUP FRONTS IN THE SINGLE-LAYER IONOSPHERE, IID 165. (a) Phase fronts and (b) group fronts.

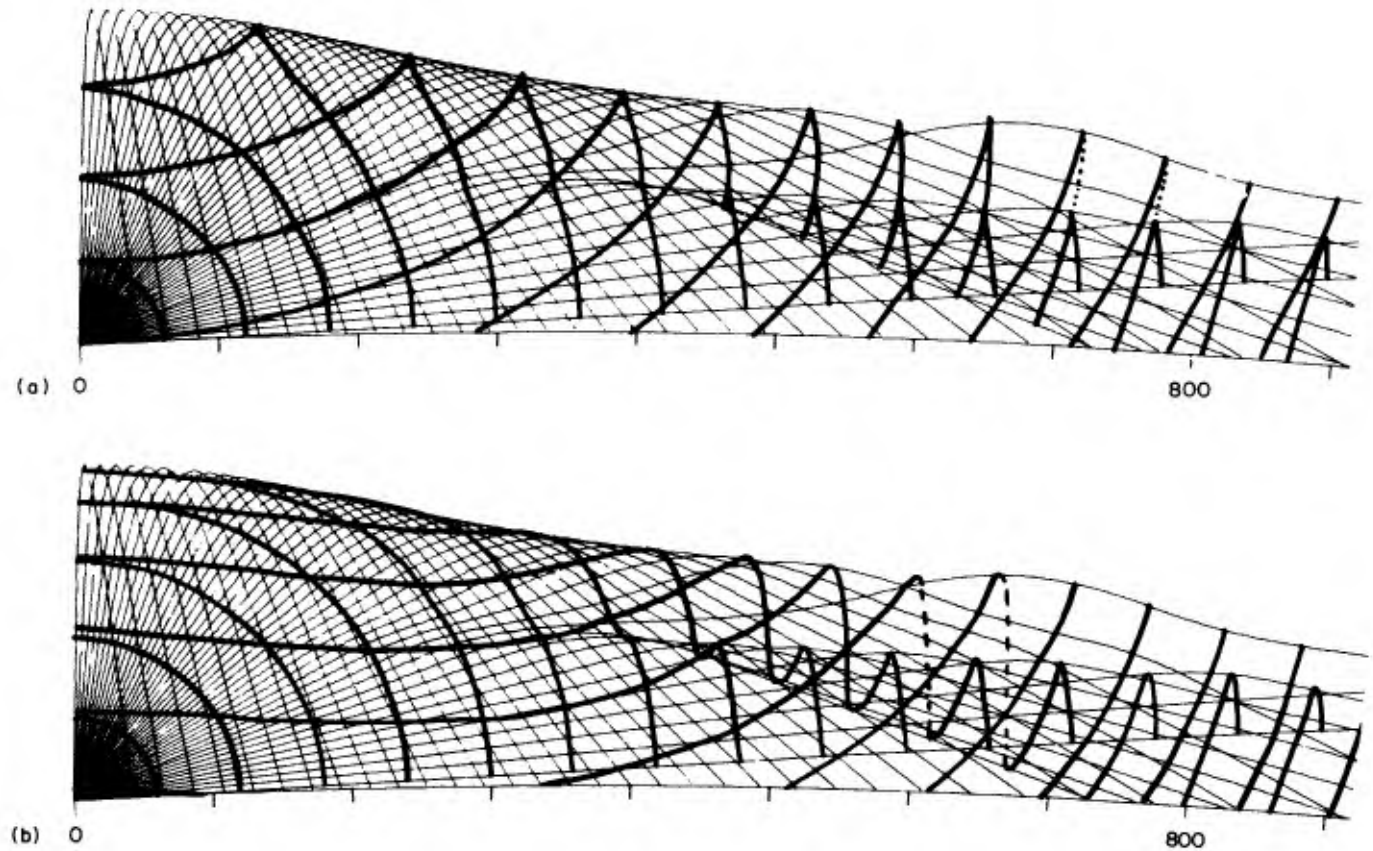


Fig. 64. 8 MHz PHASE AND GROUP FRONTS IN THE TWO-LAYER IONOSPHERE, IID 166. (a) Phase fronts and (b) group fronts.

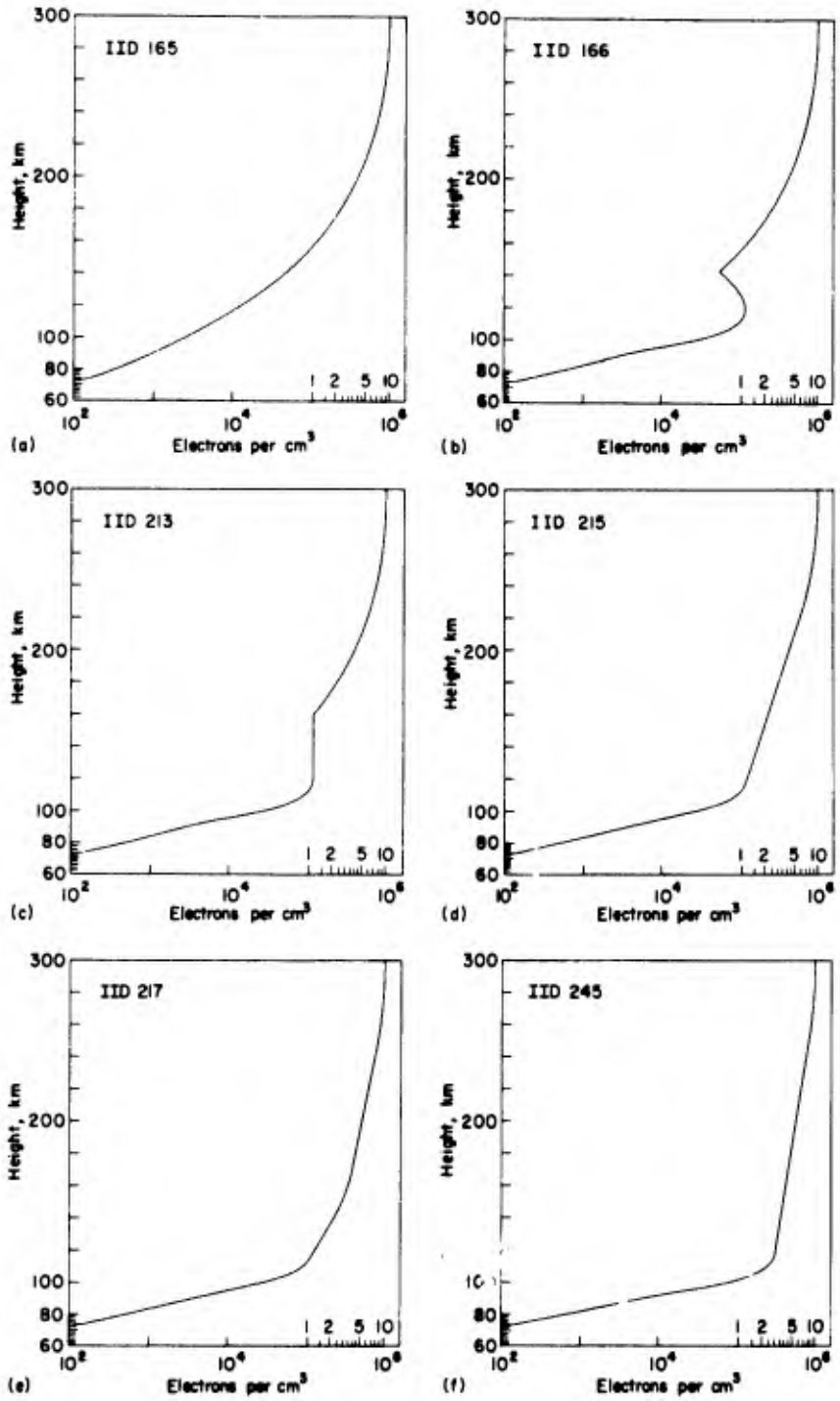


Fig. 65. IONOSPHERIC MODELS USED TO PRODUCE MANY OF THE PRECEDING RAY PLOTS. (a) IID 165, (b) IID 166, (c) IID 213, (d) IID 215, (e) IID 217, and (f) IID 245.

REFERENCES

1. J. M. Kelso, Ray tracing in the ionosphere, Radio Sci. 3 (New Series), 1, 1968.

2. F. Kift, The propagation of HF radio waves to long distances, Proc. IEEE 107B, 1960.

3. G. F. Fooks, HF propagation program, Dept. of Sci. and Indus. Res., Radio Research Station, I.M. 32, 1962.

4. D. E. Westover and L. A. Roben, Adaptation of the Kift-Fooks ionospheric ray-tracing technique to a high-speed digital computer, TR No. 78, Radioscience Laboratory, Stanford, California, 1963.

5. D. L. Neilson, Stanford Research Institute, Menlo Park, California, private communication, 26 June 1966.

6. J. Haselgrove, Ray theory and a new method for ray tracing, Report on Conf. on Phys. of the Ionosphere, Phys. Soc. (London) 1954, p. 355.

7. J. Haselgrove, Oblique ray paths in the ionosphere, Proc. Phys. Soc., B, 70, 1957.

8. K. G. Budden and G. J. Daniell, Rays in magnetoionic theory, J. Atmos. Terres. Phys. 27, 1964.

9. K. Suchy and A. K. Paul, Hamilton's equations of geometric optics, Technical Note BN-424, Institute for Fluid Dynamics & Applied Mathematics, University of Maryland, College Park, Maryland, 1965.

10. J. E. Titheridge, Ray-paths in the ionosphere, J. Atmos. Terres. Phys. 14, 1959.

11. K. Davies, Oblique propagation, Chapter 4 in Ionospheric Radio Propagation, Monograph 80, National Bureau of Standards, 1965.

12. H. Poeverlein, Strahlweze von radiowellen in der ionosphare, Z. angen Phys. 1, 1949.

13. H. G. Booker, Application of magnetoionic theory to radio waves incident obliquely upon horizontally stratified ionosphere, J. Geophys. Res. 54, 1949.

14. M. S. Wong, Ray tracing study of satellite-to-satellite HF propagation, Rad. Astron. Satell. Studies of the Atmosph., Air Force Cambridge Research Lab., Office of Aerospace Research, Boston, Massachusetts, 1965.

15. O. K. Garriott, Determination of ionospheric electron content and distribution from satellite observations, Part I: Theory of the analysis, J. Geophys. Res. 65, 1960.

16. T. A. Croft, Interpreting the structure of oblique ionograms, TR No. 114, Radioscience Laboratory, Stanford, California, 1966.

17. T. A. Croft, Computation of HF ground backscatter amplitude, Radio Sci. 2 (New Series), 7, 1967.

18. C. R. Gilliland, Sweep-frequency backscatter with calibrated amplitude, TR No. 111, Radioscience Laboratory, Stanford, California, 1965.

19. T. A. Croft, The influence of ionospheric irregularities on sweep-frequency backscatter, J. Atmos. Terres. Phys. 30, 1968.

20. W. J. G. Beynon and E. S. Owen Jones, Some medium latitude radio wave absorption studies, J. Atmos. Terres. Phys. 27, 1965.

21. H. Poeverlein, Field strength near the skip distance, TR No. 54-104, Propagation Laboratory, Air Force Cambridge Research Center, Cambridge, Massachusetts, 1955.

22. H. Bremmer, Terrestrial Radio Waves, American Elsevier Publ. Co., New York, 1949.

23. T. A. Croft and H. Hoogasian, Exact ray calculations in a quasi-parabolic ionosphere with no magnetic field, Radio Sci. 3 (New Series), 1, 1968.

24. D. E. Westover, Exact ray-path solutions in a quasi-linear ionosphere, Radio Sci. 3 (New Series), 1, 1968.

25. R. M. Jones, A three-dimensional ray-tracing computer program, Radio Sci. 3 (New Series), 1, 1968.

26. E. Appleton and W. J. G. Beynon, The application of ionospheric data to radio communication problems, Proc. Phys. Soc. 52, Part 1, 1940.

27. D. L. Nielson, Ray-path equations for an ionized layer with a horizontal gradient, Radio Sci. 3 (New Series), 1, 1968.

28. K. Davies, A nomenclature for oblique ionospheric soundings and ray tracing, Radio Sci. 2 (New Series), 11, 1967.

29. Ya. L. Al'pert, The refraction and doppler shift of radio waves emanating from an artificial earth satellite in a three-dimensional nonhomogeneous ionosphere, Geomagnetism and Aeronomy III, 4, 1963.

30. W. V. Lovitt, Linear Integral Equations, Dover Publications Inc., New York, 1950.

31. T. A. Croft and L. Gregory, A fast, versatile ray-tracing program for IBM 7090 digital computers, TR No. 82, Radioscience Laboratory, Stanford, California, 1963.

32. T. A. Croft, Ionospheric structure determination from vertical-angle measurements, TR No. 115, Radioscience Laboratory, Stanford, California, 1966.

33. T. J. Lemanski, An evaluation of the ITSA 3-D ray-trace program, Radio Sci. 3, (New Series), 1, 1968.

34. T. A. Croft, HF radio focusing caused by the electron distribution between ionospheric layers, J. Geophys. Res. 72, 9, 1967.

ONR-ARPA U-SERIES DISTRIBUTION LIST

NAVY

Chief of Naval Research
Department of the Navy
Washington, D.C. 20360
2 Attn: Code 418

Director
Naval Research Laboratory
Washington, D.C. 20390
1 Attn: Code 5320 (E. Zettle)
1 Attn: Code 5320 (J.M. Headrick)
1 Attn: Code 5432C (F.A. Polinghorn)
1 Attn: Code 2027
1 Attn: Code 5400 (L. Wetzel)

Chief of Naval Operations
Department of the Navy
The Pentagon
Washington, D.C. 20350
1 Attn: OP-07TE
1 Attn: OP-723E

Commander Naval Missile Center
Point Mugu, California 93041
1 Attn: Code NO3022

Commander
Naval Weapons Center
China Lake, California 93555
1 Attn: Code 4025 (R.S. Hughes)

Commander
Naval Electronics Laboratory Center
San Diego, California 92152
1 Attn: Mr. H.J. Wirth
1 Attn: Library

Commander
Naval Weapons Center
Corona Laboratories
Corona, California 91720
1 Attn: Mr. V.E. Hildebrand

Director
Office of Naval Resear[illegible]ranch Office
495 Summer Street
Boston, Massachusetts 02210
1 Attn: Mr. Stan Curley

AIR FORCE

Headquarters, USAF
The Pentagon
Washington, D.C. 20330
1 Attn: AFNICAD (MAJ Nyquist)
1 Attn: AFRDDF

Headquarters, USAF
Office of Assistant Chief
of Staff, Intelligence
Washington, D.C. 20330
1 Attn: AFNICAA

Commander
Rome Air Development Center
Research & Technology Div.
Griffiss AFB, New York 13442
1 Attn: EMASC (S. DiGennaro)
1 Attn: EMAES (MAJ Wipperman)
1 Attn: EMASR (V.J. Coyne)
1 Attn: EMASA

Headquarters
Air Force Systems Command
Foreign Technology Division
Wright-Patterson AFB
Ohio 45433
1 Attn: TDDBP (Mr. Zabetakis)
1 Attn: TDEED (W.L. Picklesimer)
1 Attn: TDCES
1 Attn: TDCE (M.S.J. Grabener)

Headquarters
Air Force Systems Command
Research & Technology Division
Bolling Air Force Base
Washington, D.C. 20332
1 Attn: RTTC

Headquarters
USAF Security Service (OSA)
San Antonio, Texas 78241
1 Attn: ODC-R (W.L. Anderson)

AIR FORCE (Cont.)

Headquarters
Air Defense Command
Ent AFB
Colorado Springs, Colo. 80912
1 Attn: NPSD-A
1 Attn: ADLPC-2A (LCOL R.J. Kaminski)
1 Attn: ADOAC-ER
1 Attn: NELC-AP

Electronics Systems Division (ESSL)
L.G. Hanscom Field
Bedford, Massachusetts 01731
1 Attn: Code 440L

Headquarters SAC (OAI)
Offutt Air Force Base
Omaha, Nebraska 68113
1 Attn: Mrs. E. G. Andrews

Headquarters, AFCRL
L.G. Hanscom Field
Bedford, Massachusetts 01731
1 Attn: CRUI
1 CRUP (Dr. G.J. Gassman)

Headquarters
U.S. Air Force
Air Force Western Test Range
Vandenberg AFB, Calif. 93437
1 Attn: WTGT (Mr. Stanley Radom)

Headquarters, USAF AFTAC
Washington, D.C. 20333
1 Attn: TD-3

Headquarters
Air Weather Service
Scott AFB, Illinois 62265
1 Attn: MAJ T.D. Damon (AWVDC)

ARMY

Office of the Assistant Chief
of Staff for Intelligence
Department of the Army
The Pentagon, Room 2B 457
Washington, D.C. 20310
1 Attn: Mr. Joseph Grady

U. S. Army SLAG
The Pentagon, Room 1B 657
Washington, D.C. 20310
1 Attn: Mr. N.R. Garofalo

Chief
Army Security Agency
Arlington Hall Station
Arlington, Virginia 22212
1 Attn: Mr. R.R. Neill
1 Attn: IAOPS-O(SA)

Commanding Officer
Army Security Agency
Processing Center
Vint Hill Farms Station
Warrenton, Virginia 22186
1 Attn: LT Alan Bagully
1 Attn: Technical Library

Commander
U.S. Army
Electronics Warfare Lab
Mt. View Office, USAEC
P.O. Box 205
Mt. View, California 94040
1 Attn: Mr. Joseph Bert

U.S. Army Foreign Science &
Technology Center
Munitions Building
Washington, D.C. 20315
1 Attn: Communications &
Electronics Division

Commanding General
U.S. Army Missile Command
Redstone Arsenal, Alabama 35809
1 Attn: AMSMI-RES

DEPARTMENT OF DEFENSE

Director
Advanced Research Projects Agency
The Pentagon
Washington, D.C. 20301
1 Attn: Mr. Alvin Van Every

DEPARTMENT OF DEFENSE (Cont.)

Office of the Assistant Director
Intelligence & Reconnaissance
Office of the Director of Defense
Research & Engineering
The Pentagon, Room 3E 119
Washington, D.C. 20301
1 Attn: Mr. H. A. Staderman

Director
National Security Agency
Fort George G. Meade
Maryland 20755
1 Attn: R-344(Mr. C. Gandy)
1 Attn: C3-TDL

Deputy Director
Research & Technology
Office of the Director of
Defense Research & Engineering
The Pentagon, Room 3E 1030
Washington, D.C. 20301
1 Attn: Dr. C. W. Sherwin

Office of the Assistant Director
(Defense Systems)
Defense Research & Engineering
The Pentagon, Room 3D 138
Washington, D.C. 20301
1 Attn: Mr. Daniel Fink

Director
Defense Intelligence Agency
The Pentagon, Room 3B 259
Washington, D.C. 20301
1 Attn: DIACO-4
1 Attn: DIAST-2B

Director
Weapons Systems Evaluation Group
Office of the Director of Defense,
Research & Engineering
1 Washington, D.C. 20301

Defense Documentation Center
Cameron Station
Alexandria, Virginia 22314
20 Attn: Document Control

National Aeronautics & Space
Administration
Ames Research Center
Moffett Field, California 94035
1 Attn: Dr. Kwok-Long Chan
1 Attn: Mr. Lawrence Colin

OTHERS

ITT Electro-Physics Laboratories,
Incorporated
3355 52nd Avenue
Hyattsville, Maryland 20781
1 Attn: Mr. W.T. Whelan

Institute for Defense Analyses
400 Army-Navy Drive
Arlington, Virginia 22202
1 Attn: Mr. Charles Lerch

MITRE Corporation
E Building, Room 353
Bedford, Massachusetts 01730
1 Attn: Mr. W.A. Whitcraft, Jr.
1 Attn: Mr. Bill Talley

RAND Corporation
1700 Main Street
Santa Monica, Calif. 90406
1 Attn: Dr. Cullen Crain
1 Attn: Library

Raytheon Company
Spencer Laboratory
2 Wayside Road
Burlington, Massachusetts 01803
1 Attn: Mr. L.C. Edwards

Stanford Research Institute
Menlo Park, California 94025
1 Attn: Dr. David Johnson

Sylvania Electronics Systems
Electronics Defense Laboratory
P.O. Box 205
Mt. View, California 94094
1 Attn: Dr. James Burke

OTHERS (Cont.)

Mr. Thurston B. Soisson
Box 8164 SW Station
1 Washington, D.C. 20024

Astrophysics Research Corporation
10889 Wilshire Boulevard
Los Angeles, California 90024
1 Attn: Dr. Alfred Reifman

Institute of Science & Technology
The University of Michigan
P.O. Box 618
Ann Arbor, Michigan 48105
1 Attn: BAMIRAC Library

Bendix Corporation
Bendix Radio Division
Baltimore, Maryland 21204
1 Attn: Mr. John Martin

AVCO Systems Division
Lowell Industrial Park
Lowell, Massachusetts 01851
1 Attn: Mr. Sidney M. Bennett

U.S. Department of Commerce
ITSA-ESSA
Boulder, Colorado 80302
1 Attn: Mr. William Utlaut
1 Attn: Mr. L.H. Tveten
1 Attn: Mr. W.A. Klemperer,
Div. 530, ESSA

Page Communications, Inc
3300 Whitehaven Street NW
1 Washington, D.C. 20008
1 Attn: Mr. David Falest, III

Massachusetts Institute of Technology
Lincoln Laboratory
P.O. Box 73
Lexington, Massachusetts 02173
1 Attn: Dr. J.H. Chisholm

Massachusetts Institute of Technology
Center for Space Research
Building 33-109
Cambridge, Massachusetts 02138
1 Attn: Dr. J.V. Harrington

Institute for Defense Analyses
100 Prospect Avenue
Princeton, New Jersey 08540
1 Attn: Dr. Edward Frieman

University of California
Mathematics Department
Berkeley, California 94720
1 Attn: Dr. E.J. Pinney

University of California
Electronics Research Laboratory
Berkeley, California 94720
1 Attn: Prof. D.J. Angelakos

Battelle-Defender
Battelle Memorial Institute
505 King Avenue
1 Columbus, Ohio 43201

HRB-Singer, Incorporated
Science Park
P.O. Box 60
State College, Pennsylvania 16801
1 Attn: Library

Pickard & Burns, Research Dept.
103 Fourth Avenue
Waltham, Massachusetts 02154
1 Attn: Dr. J.C. Williams

Sylvania Electronic Systems
Applied Research Laboratory
40 Sylvan Road
Waltham, Massachusetts 02154
1 Attn: Library

OTHERS (Cont.)

Department of Electrical Engineering
Radiolocation Research Laboratory
University of Illinois
Urbana, Illinois 61803
1 Attn: 311 EERL (Mr. D.G. Detert)

Arecibo Ionospheric Observatory
Box 995
Arecibo, Puerto Rico 00613
1 Attn: Librarian

The University of Texas
Electrical Engineering Research Laboratory
Route 4, Box 189
Austin, Texas 78756
1 Attn: Mr. C.W. Tolbert

Rice University
Fondren Library
P.O. Box 1892
1 Houston, Texas 77001

Purdue University
Library
1 West Lafayette, Indiana 47906

Telcom, Incorporated
8027 Leesburg Pike
McLean, Va. 22101
1 Attn: Mr. J.D. Ahlgren,
Vice President

IBM
Armonk, New York 10504
1 Attn: S.W. Woolven
(Info Retrieval Specialist)

General Telephone & Electronics
Laboratories, Inc.
208-20 Willets Point Boulevard
Bayside, New York 11360
1 Attn: Mr. George H. Kiesewetter

Raytheon Company
Research Division Library
28 Seyon Street
1 Waltham, Massachusetts 02154

Barry Research
934 E. Meadow Drive
Palo Alto, California 94303
1 Attn: Ronald L. Murphy

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)*

Stanford Electronics Laboratories
Stanford University, Stanford, California

2a. REPORT SECURITY CLASSIFICATION

Unclassified

2b. GROUP

3. REPORT TITLE

METHODS AND APPLICATIONS OF COMPUTER RAYTRACING

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

Technical Report

5. AUTHOR(S) *(First name, middle initial, last name)*

Thomas A. Croft

6. REPORT DATE	7a. TOTAL NO OF PAGES	7b. NO OF REFS
January 1969	157	34

8a. CONTRACT OR GRANT NO	9a. ORIGINATOR'S REPORT NUMBER(S)
Nonr-225(64)	SU-SEL-69-007
b. PROJECT NO NR 088-019	TR No. 112
c.	9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*
d.	

10. DISTRIBUTION STATEMENT

This document is subject to special export controls and each transmittal to foreign governments or foreign nationals may be made only with prior approval of the Office of Naval Research, Field Projects Programs, Washington, D. C. 20360.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Office of Naval Research ARPA Order No. 196

13. ABSTRACT

For about five years, considerable effort was devoted at this laboratory to the exploitation of the digital computer as a means for simulating ionospheric radio propagation and associated signal processing. Stress was placed on simulation of oblique propagation at HF by the most practical means, engineering the programs to achieve economical operation. The computer processes consist mainly of ionospheric structural analysis and coding, raytracing, and the subsequent use of ray trajectory data to simulate the operation of complete ionospheric radio systems.

This report summarizes the conclusions which have been drawn from the work, including much technical data, some programming details, and hopefully some insight into the optimum methods for doing this work. Also, there is a considerable amount of information inherently contained in these data which bear on the understanding of the propagation simulated. Many applications of raytracing are discussed, and many useful programming or plotting techniques are described, some for the first time. Finally, a number of plotted examples are given to show what can be achieved by these methods and to illustrate selected aspects of ionospheric propagation.

DD FORM 1 NOV 65 1473

UNCLASSIFIED

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
IONOSPHERE						
RAYTRACING						
COMPUTER SIMULATION						
RADIO, HF						